2011

Environmental analyses and candidate gene expression of ms9 male-sterility locus and computational identification of candidate genes for male-sterility loci in soybean

Christiana Wiebbecke
Iowa State University

Environmental analyses and candidate gene expression of *ms*9 male-sterility locus and computational identification of candidate genes for male-sterility loci in soybean

by

Christiana Elaine Wiebbecke

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Silvia R. Cianzio, Co-major Professor
Michelle A. Graham, Co-major Professor
Reid G. Palmer, Co-major Professor
Thomas Lübberstedt
Daniel J. Nordman
Andrew D. Nickell

Iowa State University

Ames, Iowa

2011

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

APPENDIX A: LIST OF SUPPLEMENTAL TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major professors Silvia R. Cianzio, Michelle A. Graham, and Reid G. Palmer for all their support, encouragement, and most of all their willingness to challenge me to think beyond my perceptions of what is possible. Your endless mentoring has provided a solid foundation for me to develop as I become a soybean breeder. Second, I would like to thank my manager Andy Nickell for his continued support to 'get the true graduate student experience' while at Iowa State, and our discussions encompassing the art and science of soybean variety development. I would also like to express my appreciation to Monsanto Soybean Breeding for providing me the opportunity to return to graduate school at Iowa State while maintaining my career and work relationships.

I would also like to thank my family and friends for all of their support of my aspiration to become a soybean breeder and gain a Ph.D. Without your confidence and encouragement this endeavor would have been meaningless. Most of I would like to thank my husband, Jay. You are truly my champion, and I love you.

ABSTRACT

Rapid growth rate of the human population is driving the increased demand for cultivated soybean [*Glycine max* (L.) Merrill], a major source of protein and oil in global food, feed, and biofuel production. Increasing yield at a pace faster than traditional breeding methods presently achieve has become a need and future commitment. In soybean, a self-pollinated crop, a possibility to increase yield may result from the development of hybrid cultivars. Researchers have been interested in floral development and pollination, as these factors have a direct impact on maintaining and enhancing crop yields. Soybean breeders also need to understand phenotypic differences in male-sterility loci that could be manipulated for use in hybrid seed production. Among them, the *ms9* locus is particularly interesting because it has high insect pollinator attraction, which facilitates its potential use in hybrid seed production. Recent release of multiple genomic tools for soybean, new bioinformatic approaches can be used to identify candidate male-sterility genes and compare their function. This dissertation is organized in chapters dealing with different aspects related to the use of male sterility.

In Chapter 2, we determined that *ms9* is a thermosensitive male-sterility locus, influenced by day temperature. Bioinformatic analyses in Chapter 3 identified four homeologs of male-sterility genes from Arabidopsis and rice as candidate genes for *ms3*, *msp*, and *ms2* loci. A total of 23 candidate genes for the *ms9*, *msp*, *ms3*, and *ms2* loci were identified. In Chapter 4, utilizing a similar bioinformatic approach targeting the *ms9* locus, we prioritized a list of nine candidate genes for analyses using quantitative real-time polymerase chain reaction (qRT-PCR). Two genes, a small nuclear ribonucleoprotein (snRNP, Glyma03g30880) and an auxin-indole-3-acetic acid (AUX/IAA) responsive gene

(Glyma03g31530), had increased expression in fertile and in male-sterile floral phenotypes as day temperature increased from 30 to 35 °C.

Our results will facilitate protocol development to utilize *ms9* as the source of male sterility for soybean breeding and hybrid seed production programs. The bioinformatic approaches used, identified for the first time, candidate male-sterility genes in soybean, which allow future elucidation of functional aspects of these candidate male-sterility genes. Finally, polymorphic markers will be designed once the male-sterility gene is confirmed. Marker assisted selection (MAS) programs can be used to identify male-sterile, female-fertile plants prior to flowering. This approach would ultimately improve efficiency of identifying plants for use as female (pod) parent in hybrid seed production.

CHAPTER 1: GENERAL INTRODUCTION

The cultivated soybean [*Glycine max* (L.) Merrill] is one of the main oilseed crops in the world (Economic Research Service, 2010). The United States is the world's leading soybean producer (35%), planting 30.33 million hectares averaging 2,811 $kg \cdot ha^{-1}$ in 2011 (Agricultural Statistics Board, 2011) with Brazil second in production at 27% (SoyStats, 2011). Soybeans are the largest source of protein feed and second largest source of vegetable oil in the world (Economic Research Service, 2010).

Breeding efforts and cultivar development have focused mainly on increasing soybean yield through selection of high-yielding genotypes within heterogeneous families that are the result of artificial cross-pollination. Soybean is a naturally self-pollinated crop, and commercial cultivars are highly inbred genotypes. Average yield increases of 15.1 $kg \cdot ha^{-1}$, or 0.6%, per year since the 1920s have been reported (Specht and Williams, 1984). Methods to achieve higher yields, other than conventional soybean cultivar development practices, need to be identified. One possibility is the utilization of hybrid soybeans for commercial production.

Exploitation of hybrid vigor in naturally self-pollinated crops has achieved vyield increases in commercially important commodities such as rice (Li and Yuan, 2000; Virmani et al., 2003), pigeonpea (Saxena et al., 2010), and cotton (McCarty et al., 2004a, 2004b). In soybean this is not a production practice, however, it could become a feasible approach. Soybean is a highly autogamous species with perfect flowers and will require considerable effort to economically produce hybrids for commercial production. Palmer et al. (2001) listed requirements, which must be satisfied for the successful development of commercial hybrid soybean cultivars: 1) parental combinations that produce heterosis levels superior to

the best pure-line cultivars, 2) a stable and uniform population of male-sterile, female-fertile soybean plants to serve as the female parent in hybrid production, 3) a selection system to obtain 100% female (pod parent) plants that set seed normally and can be harvested mechanically, 4) an efficient pollen transfer mechanism from pollen parent to pod parent, and 5) an economic level of seed increase for seedsmen and growers that ultimately will benefit the consumer. These requirements will be discussed in relation to research we conducted in soybean to understand male-sterility loci, with reference to the *ms9* locus.

The first requirement to consider when incorporating hybrids into commercial production for a crop is the extent to which yield heterosis is expressed. For soybean, high-parent heterosis ranged from 2 to 20% in replicated field plot experiments (Palmer et al., 2001). High parent heterosis was evident between specific parental combinations among 73 hybrids, which ranged from -10% to 20%, with 36 hybrid combinations achieving a mean yield greater than the high parent (Palmer et al., 2001). This suggests that dominant genetic effects accompany the significant additive × additive epistatic effects that have been reported in soybean breeding populations (Burton, 1987).

The second requirement for hybrid seed production is the identification of stable and uniform male-sterile, female-fertile plants to serve as the female parent. Complete sterility in soybean first was observed in progeny of Manchu (Owen, 1928). Sterile plants did not set seed and maintained green leaves instead of progressing through normal reproductive development from stages R1 through R8 as defined by Fehr and Caviness (1977). In addition to the complete sterile phenotype described above, partial-female sterility (Pereira et al., 1997), partial-male sterility (Caviness et al., 1970; Stelly and Palmer, 1980), and male sterility (Horner and Palmer, 1995; Palmer et al., 2001) also have been observed in soybean.

Prior evidence suggests soybean lines expressing male sterility may be environmentally sensitive (Caviness and Fagala, 1973; Stelly and Palmer, 1980). Environmental sensitivity in relation to stable expression of male-sterility needs to be investigated for different male-sterility loci prior to their use and incorporation into a hybrid seed production system.

For self-pollinating crops, i.e. soybean, the amount of out-crossed seed is directly related to the magnitude of attraction by insects (Palmer et al., 2009), which relates to the fourth requirement for hybrid seed production in soybean, efficient pollen transfer. Insects are typically the only means by which to transfer pollen between plants resulting in natural cross-pollination. For soybeans, the frequency of cross-pollination generally is less than 1% (Fehr, 1980). The male-sterility locus *ms9* originally was selected for our research because an evaluation of insect-mediated seed-set demonstrated, that it had the highest out-crossing capability and seed-set of ten male-sterility loci previously identified (Ortiz-Perez et al., 2006). Soybean breeders need to understand floral characteristics and insect-pollinator relations since these factors directly relate to hybrid seed production, potential hybrid yield, and yield stability (Palmer et al., 2009). An additional challenge related to efficient pollen transfer and cross-pollination to produce hybrid seed in naturally autogamous crops is, how to limit self-pollination. For this reason, understanding male-sterility loci in soybean is a crucial component for practical execution of hybrid seed production.

In using male-sterility systems for hybrid production in soybean, the classification of fertile versus male-sterile, female-fertile plants needs to be done. This is accomplished by visually examining pollen phenotype at flowering time, the R2 stage as defined by Fehr and Caviness, (1977). Flower classification causes delays in manual cross-pollination and is labor intensive. Therefore, my initial research objective was to address the third requirement

for hybrid seed production, i.e. an effective phenotypic selection method to identify male-sterile, female-fertile plants prior to flowering. The original goal was to develop molecular marker(s) that could be used in a marker assisted selection (MAS) program as a means to identify male-sterile, female-fertile plants prior to flowering, so the *ms9* gene could be used in hybrid soybean development.

With the newly developed soybean genome browser (www.soybase.org), we were able to visualize the *ms9* genomic region. BLASTN ($E < 10^{-4}$, Altschul et *al.*, 1997) analysis was used to compare the flanking simple sequence repeat (SSR) markers Satt521 and Satt237 identified by Cervantes-Martinez et al. (2007) to the soybean genome sequence (Glyma 1, Schmutz et al., 2010). This identified the intervening region on chromosome 03 (molecular linkage group, MLG N) at position 38.6 to 40.1 mega bases ($= 10^6$ base pairs, Mb). This sequence was searched for AT (15) and AAT (10) repeats that were used to develop novel SSR markers. In total, 68 primers were designed to flank the repeats in the hope that the mapping parents would have polymorphisms due to differences in repeat number. Primers were tested against the parents of Soybean Genetic Type Collection number T359H, the heterozygous (H) maintainer line for *ms9*, to identify polymorphic markers that could be used to screen the selfed progeny of T359H.

Challenges arose using pollen phenotype to classify male-sterile, female-fertile plants of selfed progeny of T359H, planted at Kunia, Hawaii, in September and October 2008. At R2 stage, no obvious differences in pollen characteristics were noted to distinguish between fertile and male-sterile, female-fertile plants at Hawaii. At harvest, the observed phenotypic segregation ratios were different from the expected ratio for a single male-sterility gene with recessive *ms9* allele. To better understand the phenotypic variation observed, we grew

T359H at Isabela, Puerto Rico in April 2009. Similar challenges were faced in identifying male-sterile, female-fertile plants when pollen phenotypic data were collected at stage R2. Only two of 151 plants that were examined, were classified as male-sterile pollen phenotype. At harvest, fertile vs. male-sterile, female-fertile plants were indistinguishable using seed-set numbers. Molecular screening of the progeny and parents was conducted with the newly designed markers for the *ms9* region. However, we were unable to confidently associate the SSR markers with the male-sterile phenotype for *ms9*.

In a separate experiment planted in Huxley, Iowa in May 2009, it was possible to identify male-sterile, female-fertile plants from the selfed progeny of T359H. The preliminary research we conducted in Hawaii, Puerto Rico, and Iowa indicated that differences in environmental conditions altered the phenotypic expression of *ms9*, increasing the difficulty of visual identification of male sterility using pollen shed at anthesis or seed-set at maturity. On the basis of these observations, our research objectives were modified. It was decided that improving the understanding of environmental influence on male sterility at the *ms9* locus was required. Research was planned accordingly, and both Chapters 2 and 4 of the dissertation were the products of this effort. Concurrently, we shifted our approach to identify candidate genes for mapped male-sterility loci in soybean and compare genomic differences between environmentally sensitive (*ms9*, *ms8*, *msp*, and *ms3*) mutants and environmentally stable male-sterility locus (*ms2*), as a part of Chapter 3. Our final research objective, Chapter 4 in the dissertation, presents our molecular approach to identify candidate genes for the *ms9* locus, for expression analysis. Results from the redefined objectives became the subject matter of additional research.

DISSERTATION ORGANIZATION

The dissertation is organized in 3 chapters, corresponding to each of the objectives. Each chapter will be submitted to appropriate journals: Chapter 2 to Crop Science, Chapter 3 to Plant Genome, and Chapter 4 to Theoretical and Applied Genetics. Additional sections in the dissertation include: general introduction, dissertation organization, objectives of each chapter, a literature review-summary, and a conclusion section.

OBJECTIVES

Chapters for publication have been written and presented as journal papers with the format of the journal to which they will be submitted.

Chapter 2

Our objectives were to characterize the phenotypic variation of pollen phenotypes and seed-set among selfed progeny of T359H, in which *ms9 ms9* plants are male-sterile, female-fertile (Palmer, 2000). This would allow us to determine if the *ms9* locus conferring male sterility in soybean was environmentally sensitive in its phenotypic expression. If it was, objectives were developed to determine which types of environments could be used for hybrid seed production and/or male-sterile line maintenance.

Chapter 3

Our objectives in this analysis were to use computational methods, to identify candidate genes for male sterility. In addition, we wanted to determine if the same transcription factor regulated the expression of all environmentally sensitive male-sterility loci. While we could not track transcription factors directly, we could determine if transcription factor binding

sites (TFBS) were common across environmentally sensitive male-sterility loci or unique for each male-sterility locus.

Chapter 4

Our objectives were to identify *ms9* candidate genes and analyze the RNA expression patterns in floral tissue of fertile and male-sterile, female-fertile plants across different controlled environments. Candidates not expressed during floral development were eliminated from further analysis. This approach enabled us to detect genes with differential expression in fertile and male-sterile, female-fertile plants across day and night temperatures. In addition, we wanted to identify differentially expressed genes between fertile and male-sterile, female-fertile plants. This approach will prioritize the list of candidate genes involved in male sterility at the *ms9* locus.

GENERAL LITERATURE REVIEW – SUMMARY APPROACH

Since sterility first was observed in soybean over 80 years ago, there have been numerous studies about the genetics of mutant nuclear genes with Mendelian inheritance, which affect male cell and organ development. The genes collectively have been designated as *ms* genes (Horner and Palmer, 1995). To date there are 11 distinct, independent male-sterility loci identified in soybean (*ms1* thru *ms9* and *msp*, Palmer et al., 2001) and *ms MOS* (Midwest Oil Seeds, Jin et al., 1997). Five loci have been mapped, *ms9* (Cervantes-Martinez et al., 2007), *ms3* (Cervantes-Martinez et al., 2009), *ms8*, *msp* (Frasch et al., 2010), and *ms2* (Cervantes-Martinez et al., 2007). All arose as spontaneous mutations with the exception of *ms5*, which was induced by mutagenesis (Buss, 1983; Graybosch and Palmer, 1988). In addition, there are multiple independent mutation events for *ms1* (7 alleles), *ms2* (2 alleles), *ms3* (3 alleles), *ms4* (2 alleles), and *ms6* (2 alleles) (Graybosch and Palmer, 1988; Palmer et al., 2001). Homozygous recessive mutations for any one of these loci result in male sterility.

Typically male-sterility genes are recessive and expressed in sporophytic tissues. Cytological analyses have been done on all male-sterility loci except *ms5*, *ms7*, *ms8*, and *ms9*. These analyses revealed that male-sterility genes have abnormal microsporogenesis at the following stages: degeneration at any stage (*msp*), failure of cytokinesis leading to tapetal degeneration (*ms1* and *ms4*), degeneration of tetrads (*ms2* and *ms6*), degeneration of microspores (*ms3*), and tapetal and/or parietal abnormalities (*ms MOS*) (Palmer et al., 2001). Furthermore, cytological studies of male-sterile mutants determined that their phenotypic expression is subdivided into three categories: 1) sporogenous, indicating stamen development is normal, but abnormalities occur during microgametogenesis, 2) structural, indicating the stamen is absent or abnormal (Johns et al., 1981), and 3) functional, indicating

pollen is not produced or does not reach the stigma (Horner and Palmer, 1995). However, molecular mechanisms controlling male sterility in soybean are just at the beginning to be understood.

Male sterility results in a floral phenotype. Therefore, one can assume that the genes involved would be expressed in floral tissues and be involved in floral development. Genetic and molecular studies have found intricate gene interactions involved in anther morphology and pollen grain differentiation (Goldberg et al., 1993). In model crop systems like Arabidopsis and rice, male-sterile mutant screenings have identified genes involved in establishment of anther cell layers, archesporial specification, tapetal development, pollen wall formation, exine formation and patterning, programmed cell death (PCD) of the tapetum, hormonal regulation of pollen development, and dehiscence (Scott et al., 2004; Wilson and Zhang, 2009; Wilson et al., 2011). A summary of genes involved in anther or pollen development, putative annotations, known mutant phenotypes, and their respective literature citations are presented in Table 1. In addition, publications by Chen et al. (2007, 2009), Woo et al. (2008), Yang et al. (2009), Tang et al. (2011), Sakata et al. (2010), and Endo et al. (2009) identified genes or candidate genes for thermosensitive genic-male sterility (TGMS).

Novel approaches facilitated research on male-sterile systems and identification of candidate genes. With the recent release of the soybean genome sequence (Glyma 1, Schmutz et al., 2010) and the soybean genome browser (www.soybase.org), researchers can visualize regions of the genome for traits of interest. Sequences of each gene in the defined region can be downloaded and BLASTX comparison analyses (Altschul et al., 1997) can be used to annotate the putative function of the genes. Another novel resource in soybean is the

recently released RNA-Seq atlas (Severin et al., 2010) that allows researchers to assess expression of candidate genes in 14 different tissue types including root, nodule, young leaf, flower, pod, and seed with several different time points during pod and seed development. These data allow researchers to identify candidate genes expressed in specific tissues, developmental cycles, or constitutively expressed throughout development for traits of interest. Finally, using Clover analysis (Frith et al., 2004) to identify overrepresented transcription factor binding sites (TFBS) from the 126 plant transcription factors in the curated TRANSFAC database (Matys et al., 2006), aids in identification of elements involved in regulating gene expression. Release of these tools enables bioinformatic approaches to identify candidate male-sterility genes and the transcription factors regulating their expression.

REFERENCES

Aarts, M.G.M., R. Hodge, K. Kalantidis, D. Florack, Z.A. Wilson, B.J. Mulligan, W.J. Stiekema, R. Scott, and A. Pereira. 1997. The Arabidopsis MALE STERILITY2 protein shares similarity with reductases in elongation/condensation complexes. Plant J. 12:615-623.

Agricultural Statistics Board. 2011. Crop Production. National Agriculture Statistics Service (NASS), USDA. September 2011.

Albrecht, C., E. Russinova, V. Hecht, E. Baaijens, and S. de Vries. 2005. The *Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1* and *2* control male sporogenesis. Plant Cell 17:3337-3349.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Alves-Ferreira, M., F. Wellmer, A. Banhara, V. Kumar, J.L. Riechmann, and E.M. Meyerowitz. 2007. Global expression profiling applied to the analysis of Arabidopsis stamen development. Plant Physiol. 145:747-762.

Ariizumi, T., T. Kawanabe, K. Hatakeyama, S. Sato, T. Kato, S. Tabata, and K. Toriyama. 2008. Ultrastructural characterization of exine development of the *transient defective exine 1* mutant suggests the existence of a factor involved in constructing reticulate exine architecture from sporopollenin aggregates. Plant Cell Physiol. 49:58-67.

Ariizumi, T., K. Hatakeyama, K. Hinata, R. Inatsugi, I. Nishida, S. Sato, T. Kato, S. Tabata, and K. Toriyama. 2004. Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. Plant J. 39:170-181.

Ariizumi, T., K. Hatakeyama, K. Hinata, S. Sato, T. Kato, S. Tabata, and K. Toriyama. 2003. A novel male-sterile mutant of *Arabidopsis thaliana*, *faceless pollen-1*, produces pollen with a smooth surface and an acetolysis-sensitive exine. Plant Mol. Biol. 53:107-116.

Burton, J.W. 1987. Quantitative genetics: relevance to soybean breeding, p. 211-247. *In* J.R. Wilcox. (ed.) Soybeans: Improvement, production and uses. 2nd ed. Agronomy.

Buss, G.R. 1983. Inheritance of a male-sterile mutant from irradiated Essex soybean. Soybean Genet. Newsl. 10:104-108.

Canales, C., A.M. Bhatt, R. Scott, and H. Dickinson. 2002. *EXS*, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr. Biol. 12:1718-1727.

Caviness, C.E., and B.L. Fagala. 1973. Influence of temperature on a partially male-sterile soybean strain. Crop Sci. 13:503-504.

Caviness, C.E., H.J. Walters, and D.L. Johnson. 1970. A partially male-sterile strain of soybean. Crop Sci. 10:107-108.

Cecchetti, V., M.M. Altamura, G. Falasca, P. Constantino, and M. Cardarelli. 2008. Auxin regulates anther dehiscence, pollen maturation, and filament elongation. Plant Cell 20:1760-1774.

Cervantes-Martinez, I., M. Xu, L. Zhang, Z. Huang, K.K. Kato, H.T. Horner, and R.G. Palmer. 2007. Molecular mapping of male-sterility loci *ms2* and *ms9* in soybean. Crop Sci. 47:374-379.

Chen, R., Y. Pan, Y. Wang, L. Zhu, and G. He. 2009. Temperature-sensitive splicing is an important molecular regulation mechanism of thermosensitive genic male sterility in rice. Chinese Sci. Bull. 54:2354-2362.

Chen, R., X. Zhao, Z. Shao, Z. Wei, Y. Wang, L. Zhu, J. Zhao, M. Sun, R. He, and G. He. 2007. Rice *UDP-Glucose Pyrophosphorylase1* is essential for pollen callose deposition and its cosupression results in a new type of thermosensitive gene male sterility. Plant Cell 19:847-861.

Cheng, Y., X. Dia, and Y. Zhao. 2006. Auxin biosynthesis by the *YUCCA flavin monoxygenases* controls the formation of floral organs and vascular tissues in Arabidopsis. Genes Dev. 20:1790-1799.

Colcombet, J., A. Boisson-Dernier, R. Ros-Palau, C.E. Vera, and J.I. Schroeder. 2005. Arabidopsis *SOMATIC EMBRYOGENESIS RECEPTOR KINASES1* and *2* are essential for tapetum development and microspore maturation. Plant Cell 17:3350-3361.

Devoto, A., M. Nieto-Rostro, D. Xie, C. Ellis, R. Harmston, E. Patrick, J. Davis, L. Sherratt, M. Coleman, and J.G. Turner. 2002. *COI1* links jasmonate signaling and fertility to the SCF ubiquitin-ligase complex in Arabidopsis. Plant J. 32:457-466.

Dong, X., Z. Hong, M. Sivaramakrishnan, M. Mahfouz, and D.P.S. Verma. 2005. *Callose synthase* (*CalS5*) is required for exine formation during microgametogenesis and for pollen viability in Arabidopsis. Plant J. 42:315-328.

Economic Research Service. 2010. Soybean and oil crops briefing room: Overview. [http://www.ers.usda.gov/Briefing/SoybeansOilCrops/]. (verified 02Oct2011). USDA, ERS.

Endo, M., T. Tsuchiya, K. Hamada, S. Kawamura, K. Yano, M. Ohshima, A. Higashitani, M. Watanabe, and M. Kawagishi-Kobayashi. 2009. High temperatures cause male sterility in rice plants with transcriptional alterations during pollen development. Plant Cell Physiol. 50:1911-1922.

Fehr, W.R. 1980. Soybean, p. 589-599. *In* W.R. Fehr and H.H. Hadley. (ed.) Hybridization of crop plants. Am. Soc. of Agron., Inc., Madison, WI.

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa Agri. Home Econ. Expt. Stn., Iowa State Univ., Ames, IA.

Frasch, R.M., C. Weigand, P.T. Perez, R.G. Palmer, and D. Sandhu. 2011. Molecular mapping of 2 environmentally sensitive male-sterile mutants in soybean. J. Hered. 102:11-16.

Frith, M.C., Y. Fu, L. Yu, J. Chen, U. Hansen, and Z. Weng. 2004. Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res. 32:1372-1381.

Goldberg, R.B., T.P. Beals, and P.M. Sanders. 1993. Anther development: basic principles and practical applications. Plant Cell 5:1217-1229.

Graybosch, R.A., and R.G. Palmer. 1988. Male sterility in soybean – an overview. Am. J. Botany 75:144-156.

Guan, Y.F., X.Y. Huang, J. Zhu, J.F. Gao, H.X. Zhang, and Z.N. Yang. 2008. *RUPTURED POLLEN GRAIN1*, a member of the *MtN3/saliva* gene family, is crucial for exine pattern formation and cell integrity of microspores in Arabidopsis. Plant Physiol. 147:852-863.

Hird, D.L., D. Worrall, R. Hodge, S. Smartt, W. Paul, and R. Scott. 1993. The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana A6* gene displays similarity to β-1,3-glucanases. Plant J. 4:1023-1033.

Hord, C.L., Y.J. Sun, L.J. Pillitteri, K.U. Torii, H. Wang, S. Zhang, and H. Ma. 2008. Regulation of Arabidopsis early anther development by the mitogen-activated protein kinases, MPK3 and MPK6, and the ERECTA and related receptor-like kinases. Mol. Plant 1:645-658.

Hord, C.L., C. Chen, B.J. Deyoung, S.E. Clark, and H. Ma. 2006. The *BAM1/BAM2 receptor-like kinases* are important regulators of Arabidopsis early anther development. Plant Cell 18:1667-1680.

Horner, H.T., and R.G. Palmer. 1995. Mechanisms of genic male sterility. Crop Sci. 35:1527-1535.

Ishiguro, S., A. Kawai-Oda, J. Ueda, I. Nishida, and K. Okada. 2001. The *DEFECTIVE IN ANTHER DEHISCENCE1* gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in Arabidopsis. Plant Cell 13:2191-2209.

Ito, T., N. Nagata, Y. Yoshiba, M. Ohme-Takagi, H. Ma, and K. Shinozaki. 2007. Arabidopsis *MALE STERILITY1* encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell 19:3549-3562.

Ito, T., F. Wellmer, H. Yu, P. Das, N. Ito, M. Alves-Ferreira, J.L. Riechmann, and E.M. Meyerowitz. 2004. The homeotic protein AGAMOUS controls microsporogenesis by regulation of *SPOROCYTELESS*. Nature 430:356-360.

Jia, G., X. Liu, H.A. Owen, and D. Zhao. 2008. Signaling of cell fate determination by the TPD1 small protein and *EMS1 receptor kinase*. PNAS 105:2220-2225.

Jin, W., H.T. Horner, and R.G. Palmer. 1997. Genetics and cytology of a new genic male-sterile soybean [*Glycine max* (L.) Merr.]. Sex. Plant Reprod. 10:13-21.

Johns, C.W., X. Delannay, and R.G. Palmer. 1981. Structural sterility controlled by nuclear mutations in angiosperms. Nucleus 24:97-105.

Jung, K.H., M.J. Han, Y.S. Lee, Y.W. Kim, I. Hwang, M.J. Kim, Y.K. Kim, B.K. Nahm, and G. An. 2005. Rice *Undeveloped Tapetum1* is a major regulator of early tapetum development. Plant Cell 17:2705-2722.

Kaneko, M., Y. Inukai, M. Ueguchi-Tanaka, H. Itoh, T. Izawa, Y. Kobayashi, T. Hattori, A. Miyao, H. Hirochika, M. Ashikari, and M. Matsuoka. 2004. Loss-of-function mutations of the rice *GAMYB* gene impair α-amylase expression in aleurone and flower development. Plant Cell 16:33-44.

Lee, S., K.H. Jung, G. An, and Y.Y. Chung. 2004. Isolation and characterization of a rice *cysteine protease* gene, *OsCP1*, using T-DNA gene-trap system. Plant Mol. Biol. 54:755-765.

Li, J., and L. Yuan. 2000. Hybrid rice: Genetics, breeding, and seed production. Plant Breed. Rev. 17:15-158.

Li, N., D.S. Zhang, H.S. Liu, C.S. Yin, X.X. Li, W.Q. Liang, Z.Yuan, B. Xu, H.W. Chu, J. Wang, T.Q. Wen, H. Huang, D. Luo, H. Ma, and D.B. Zhang. 2006. The rice *Tapetum Degeneration Retardation* gene is required for tapetum degradation and anther development. Plant Cell 18:2999-3014.

Li, S.F., S. Iacuone, and R.W. Parish. 2007. Suppression and restoration of male fertility using a transcription factor. Plant Biotech. J. 5:297-312.

Mandaokar, A., and J. Browse. 2009. *MYB108* acts together with *MYB24* to regulate jasmonate-mediated stamen maturation in Arabidopsis. Plant Physiol. 149:851-862.

Matys, V., O.V. Kel-Margoulis, E. Fricke, I. Liebich, S.Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A.E. Kel, and E. Wingender. 2006. TRANSFAC and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34:D108-D110.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004a. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. Crop Sci. 44:1226-1230.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004b. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: II. Genetic effects and genotypic values. Crop Sci. 44:1231-1235.

Melzer, R., W. Verelst, and G. Theiβen. 2009. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. Nucleic Acids Res. 37:144-157.

Millar, A.A., and F. Gubler. 2005. The Arabidopsis *GAMYB-Like* genes, *MYB33* and *MYB65*, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell 17:705-721.

Mizuno, S., Y. Osakabe, K. Maruyama, T. Ito, K. Osakabe, T. Sato, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2007. *Receptor-like protein kinase 2* (*RPK 2*) is a novel factor controlling anther development in *Arabidopsis thaliana*. Plant J. 50:751-766.

Nagpal, P., C.M. Ellis, H. Weber, S.E. Ploense, L.S. Barkawi, T.J. Guilfoyle, G. Hagen, J.M. Alonso, J.D. Cohen, E.E. Farmer, J.R. Ecker, and J.W. Reed. 2005. *Auxin response factors ARF6* and *ARF8* promote jasmonic acid production and flower maturation. Development 132:4107-4118.

Nonomura, K.I., K. Miyoshi, M. Eiguchi, T. Suzuki, A. Miyao, H. Hirochika, and N. Kurata. 2003. The *MSP1* gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice. Plant Cell 15:1728-1739.

Ogawa, M., P. Kay, S. Wilson, and S.M. Swain. 2009. *ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1* (*ADPG1*), *ADPG2*, and *QUARTET2* are polygalacturonases required for cell separation during reproductive development in Arabidopsis. Plant Cell 21:216-233.

Ortiz-Perez, E., H.T. Horner, S.J. Hanlin, and R.G. Palmer. 2006. Insect-mediated seed-set evaluation of 21 soybean lines segregating for male sterility at 10 different loci. Euphytica 152:351-360.

Owen, F.V. 1928. A sterile character in soybeans. Plant Physiol. 3:223-226.

Palmer, R.G. 2000. Genetics of four male-sterile, female-fertile soybean mutants. Crop Sci. 40:78-83.

Palmer, R.G., P.T. Perez, E. Ortiz-Perez, F. Maalouf, and M.J. Suso. 2009. The role of crop-pollinator relationships in breeding for pollinator-friendly legumes: from a breeding perspective. Euphytica 170:35-52.

Palmer, R.G., J. Gai, H. Sun, and J.W. Burton. 2001. Production and evaluation of hybrid soybean. Plant Breed Rev. 21:263-307.

Paxson-Sowders, D.M., C.H. Dodrill, H.A. Owen, and C.A. Makaroff. 2001. DEX1, a novel plant protein, is required for exine pattern formation during pollen development in Arabidopsis. Plant Physiol. 127:1739-1749.

Pereira, T.N.S., N.R. Lersten, and R.G. Palmer. 1997. Genetic and cytological analyses of a partial-female-sterile mutant (*PS-1*) in soybean (*Glycine max*; Leguminosae). Am. J. Bot. 84:781-791.

Preston, J., J. Wheeler, J. Heazlewood, S.F. Li, and R.W. Parish. 2004. *AtMYB32* is required for normal pollen development in *Arabidopsis thaliana*. Plant J. 40:979-995.

Reyes, F., G. León, M. Donoso, F. Brandizzi, A.P.M. Weber, and A. Orellana. 2010. The nucleotide sugar transporters *AtUTr1* and *AtUTr3* are required for the incorporation of UDP-glucose into the endoplasmic reticulum, are essential for pollen development and are needed for embryo sac progress in *Arabidopsis thaliana*. Plant J. 61:423-435.

Rhee, S.Y., E. Osborne, P.D. Poindexter, and C.R. Somerville. 2003. Microspore separation in the *quartet 3* mutants of Arabidopsis is impaired by defect in a developmentally regulated

polygalacturonase required for pollen mother cell wall degradation. Plant Physiol. 133:1170-1180.

Rhee, S.Y., and C.R. Somerville. 1998. Tetrad pollen formation in quartet mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall. Plant J. 15:79-88.

Sakata, T., T. Oshino, S. Miura, M. Tomabechi, Y. Tsunaga, N. Higashitani, Y. Miyazawa, H. Takahashi, M. Watanabe, and A. Higashitani. 2010. Auxins reverse plant male sterility caused by high temperatures. PNAS 107:8569-8574.

Sanders, P.M., P.Y. Lee, C. Biesgen, J.D. Boone, T.P. Beals, E.W. Weiler, and R.B. Goldberg. 2000. The Arabidopsis *DELAYED DEHISCENCE1* gene encodes an enzyme in the jasmonic acid synthesis pathway. Plant Cell 12:1041-1061.

Saxena, K.B., R. Sultana, N. Mallikarjuna, R.K. Saxena, R.V. Kumar, S.L. Sawargaonkar, and R.K. Varshney. 2010. Male-sterility systems in pigeonpea and their role in enhancing yield. Plant Breed. 129:125-134.

Schiefthaler, U., S. Balasubramanian, P. Sieber, D. Chevalier, E. Wisman, and K. Schneitz. 1999. Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. PNAS 96:11664:11669.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Schiøtt, M., S.M. Romanowsky, L. Bækgaard, M.K. Jakobsen, M.G. Palmgren, and J.F. Harper. 2004. A plant plasma membrane Ca^{2+} pump is required for normal pollen tube growth and fertilization. PNAS 101:9502-9507.

Scott, R.J., M. Spielman, and H.G. Dickinson. 2004. Stamen structure and function. Plant Cell 16:S46-S60.

Severin, A.J., J.L. Woody, Y. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Spect, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, and R.C. Shoemaker. 2010. RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol. 10:160-176.

Sorensen, A.M., S. Kröber, U.S. Unte, P. Huijser, K. Dekker, and H. Saedler. 2003. The Arabidopsis *ABORTED MICROSPORE* (*AMS*) gene encodes a MYC class transcription factor. Plant J. 33:413-423.

SoyBase. [http://www.soybase.org].

SoyStats. 2011. A reference guide to important soybean facts and figures. [www.soystats.com/2011/]. (verified on 01Oct2011). American Soybean Association.

Specht, J.E., and J.H. Williams. 1984. Contribution of genetic technology to soybean productivity – retrospect and prospect, p.49-74. *In* W.R. Fehr (ed.) Genetic contributions to yield gains of five major crop plants. Am. Soc. of Agron., Inc., Madison, WI.

Stelly, D.M., and R.G. Palmer. 1980. A partially male-sterile mutant line of soybeans, *Glycine max* (L.) Merr.: Characterization of the *msp* phenotype variation. Euphytica 29:539-546.

Stintzi, A., and J. Browse. 2000. The Arabidopsis male-sterile mutant, *opr3*, lacks the 12-osophytodienoic acid reductase required for jasmonate synthesis. PNAS 97:10625-10630.

Suzuki, T., K. Masaoka, M. Nishi, K. Nakamura, and S. Ishiguro. 2008. Identification of *kaonashi* mutants showing abnormal pollen exine structure in *Arabidopsis thaliana*. Plant Cell Physiol. 49:1465-1477.

Tang, Z., L. Zhang, D. Yang, C. Zhao, and Y. Zheng. 2011. Cold stress contributes to aberrant cytokinesis during male meiosis I in a wheat thermosensitive genic male sterile line. Plant Cell Environ. 34:389-405.

Thorstensen, T., P.E. Grini, I.S. Mercy, V. Alm, S. Erdal., R. Aasland, and R.B. Aalen. 2008. The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor *ABORTED MICROSPORES* (*AMS*). Plant Mol. Biol. 66:47-59.

Tilly, J.J., D.W. Allen, and T. Jack. 1998. The CArG boxes in the promoter of the Arabidopsis floral organ identity gene *APETALA3* mediate diverse regulatory effects. Development 125:1647-1657.

Virmani, S.S., Z.M. Sun, T.M. Mou, A.J. Ali, and C.X. Mao. 2003. Two-line hybrid rice seed production, p. 53-64. *In* B. Hardy (ed.) Two-line hybrid rice breeding manual. International Rice Research Institute. Los Baños, Philippines.

Wan, J., A. Patel, M. Mathieu, S.Y. Kim, D. Xu, and G. Stacey. 2008. A lectin receptor-like kinase is required for pollen development in Arabidopsis. Plant Mol. Biol. 67:469-482.

Wang, A., Q. Xia, Q. Xie, R. Datla, and G. Selvaraj. 2003. The classical ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. PNAS 100:14487-14492.

Wilson, Z.A., J. Song, M. Taylor, and C. Yang. 2011. The final split: the regulation of anther dehiscence. J. Exp. Bot. 62:1633-1649.

Wilson, Z.A., and D.B. Zhang. 2009. From Arabidopsis to rice: pathways in pollen development. J. Exp. Bot. 60:1479-1492.

Wilson, Z.A., S.M. Morroll, J. Dawson, R. Swarup, and P.J. Tighe. 2001. The Arabidopsis *MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J. 28:27-39.

Woo, M.O., T.H. Ham, H.S. Ji, M.S. Choi, W. Jiang, S.H. Chu, R. Piao, J.H. Chin, J.A. Kim, B.S. Park, H.S. Seo, N.S. Jwa, S. McCouch, and H.J. Koh. 2008. Inactivation of the *UGPase1* gene causes genic male sterility and endosperm chalkiness in rice (*Oryza sativa* L.). Plant J. 54:190-204.

Xiao, X., Y. Yang, Y. Yang, J. Lin, D. Tang, and X. Liu. 2009. Comparative analysis of young panicle proteome in thermo-sensitive genic male-sterile rice Zhu-1S under sterile and fertile conditions. Biotech Lett. 31:157–161.

Xu, J., C. Yang, Z. Yuan, D. Zhang, M.Y. Gondwe, Z. Ding, W. Liang, D. Zhang, and Z.A. Wilson. 2010. The *ABORTED MICROSPORES* regulatory network is required for postmeiotic male reproductive development in *Arabidopsis thaliana*. Plant Cell 22:91-107.

Yamaguchi, T., D.Y. Lee, A. Miyao. H. Hirochika, G. An, and H.Y. Hirano. 2006. Functional diversification of the two *C-class MADS box* genes *OsMADS3* and *OsMADS56* in *Oryza sativa*. Plant Cell 18:15-28.

Yang, C., G. Vizcay-Barrena, K. Conner, and Z.A. Wilson. 2007. *MALE STERILITY1* is required for tapetal development and pollen wall biosynthesis. Plant Cell 19:3530-3548.

Yang, K.Z., C. Xia, X.L. Liu, X.Y. Dou, W. Wang, L.Q. Chen, X.Q. Zhang, L.F. Xie, L. He, X. Ma, and D. Ye. 2009. A mutation in *THERMOSENSITIVE MALE STERILITY1*, encoding a heat shock protein with DnaJ and PDI domains, leads to thermosensitive gametophytic male sterility in Arabidopsis. Plant J. 57:870-882.

Yang, S.L., L. Jiang, C.S. Puah, L.F. Xie, X.Q. Zhang, L.Q. Chen, W.C. Yang, and D. Ye. 2005. Overexpression of *TAPETUM DETERMINANT1* alters the cell fates in the Arabidopsis carpel and tapetum via genetic interaction with *EXCESS MICROSPOROCYTES1*/ *EXTRA SPOROGENOUS CELLS*. Plant Physiol. 139:186-191.

Yang, W.C., D. Ye, J. Xu, and V. Sundaresan. 1999. The *SPOROCYTELESS* gene of Arabidopsis is required for initiation of sporogenesis and encodes a nuclear protein. Genes Dev. 13:2108-2117.

Ye, Q., W. Zhu, L. Li, S. Zhang, Y. Yin, H. Ma, and X. Wang. 2010. Brassinosteroids control male fertility by regulating the expression of key genes involved in Arabidopsis anther and pollen development. PNAS 107:6100-6105.

Zhang, D., W. Liang, C. Yin, J. Zong, F. Gu, and D. Zhang. 2010. OsC6, encoding a lipid transfer protein required for postmeiotic anther development in rice. Plant Physiol. 154:149-162.

Zhang, W., Y. Sun, L. Timofejeva, C. Chen, U. Grossniklaus, and H. Ma. 2006. Regulation of Arabidopsis tapetum development and function by *DYSFUNCTIONAL TAPETUM* (*DYT1*) encoding a putative bHLH transcription factor. Development 133:3085-3095.

Zhang, Z.B., J. Zhu, J.F. Gao, C. Wang, H. Li, H. Li, H.Q. Zhang, S. Zhang, D.M. Wang, Q.X. Wang, H. Huang, H.J. Xia, and Z.N. Yang. 2007. Transcription factor *AtMYB103* is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J. 52:528-538.

Zhao, X., J. de Palma, R. Oane, R. Gamuyao, M. Luo, A. Chaudhury, P. Hervé, Q. Xue, and J. Bennett. 2008. *OsTDL1A* binds to the LRR domain of rice receptor kinase *MSP1*, and is required to limit sporocyte numbers. Plant J. 54:375-387.

Zhao, D.Z., G.F. Wang, B. Speal, and H. Ma. 2002. The *EXCESS MICROSPOROCYTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in Arabidopsis anther. Genes Dev. 16:2021:2031.

Zhu, J., H. Chen, H. Li, J.F. Gao, H. Jiang, C. Wang, Y.F. Guan, and Z.N. Yang. 2008. Defective in *Tapetal Development and Function 1* is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. 55:266-277.

Table 1. Male sterility related genes in model species Arabidopsis and rice.

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
Arabidopsis	Archesporial specification and establishment of anther cell layers				
	NZZ/ SPL	*NOZZLE/ SPOROCYTELESS*	Homology to MADS-like transcription factor (TF) acts during archesporial division on stamen and carpen development	Female and male sterile; *nzz* mutants fail to develop pollen mother cell (PMC) or megaspore mother cell; *spl* mutant blocks sporocyte and anther wall formation	Wilson and Zhang, 2009 Schiefthaler et al., 1999 Yang et al., 1999
	AG	*AGAMOUS*	MADS box TF; necessary for specification of reproductive organs (stamen and carpels) Pollen formation is induced by *AG* activating *NZZ*/*SPL* or *DAD1* regulating stamen development	*ag-1* mutant has indefinite numbers of successive whorls of sepal-petal-petal	Ito et al., 2004 Wilson et al., 2011
	AP3	*APETALLA3*	MADS box gene required for normal development of petals and stamen	Misexpression results in development of stamens in place of 4th whorl carpels	Tilly et al., 1998 Wilson et al., 2011
	PI	*PISTILLATA*	Necessary for proper development of 2nd and 3rd whorl of flower (petals and stamen)	Misexpression results in petaloid sepals in place of sepals in 1st whorl	Tilly et al., 1998 Wilson et al., 2011
	SEP1 *SEP2* *SEP3* *SEP4*	*SEPALLATA*	Class E genes, redundant to *Agamous-like* (*AGL2*, *AGL4*, *AGL9*, and *AGL3*), interacts with AP3 and PI	Simultaneous disruption of *SEP1*, *SEP2*, and *SEP3* leads to development of sepals rather than petals, stamen and carpels, disruption of all *SEP* leads to transformation of all floral organs into vegetative leaves	Melzer et al., 2009

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	DAD1	*DEFECTIVE IN ANTHER DEHISCENCE1*	Chloroplastic phospholipase A1 catalyzes initial step in jasmonic acid (JA) biosynthesis	Failure of pollen maturation, dehiscence, and flower opening due to lack of JA	Ishiguro et al., 2001
	EXS	*EXTRASPOROGENOUS CELLS*	EXS Leu-rich repeat (LRR) receptor kinase protein (RPK) regulates fate and number of archesporial cells in the anther	*exs* mutant male development ceases at tetrad stage, anthers have no tapetal or middle cell layer, and contain no pollen	Canales et al., 2002
	EMS1	*EXCESS MICROSPOROCYTES1*	*EMS1* LRR-RPK involved in differentiation of microsporocytes and tapetal cells	*ems1* male-sterile mutants produce extra meiocytes and lack tapetal cells but maintain middle cell layer	Zhao et al., 2002
	TPD TPD1	*TAPETUM DETERMINANT1*	Tapetal specification and maintenance of tapetal cell fate; ligand for *EMS1* receptor kinase	*tpd1* mutant phenotype like *exs/ems1*mutant	Yang et al., 2005 Jia et al., 2008
	BAM1 BAM2	*BARELY ANY MERISTEM*	LRR-receptor-like kinase (RLK) proteins; involved in early anther development regulating cell division and differentiation	*bam1 bam2* mutants are male sterile, fail to produce pollen, have abnormal formation of anther cell layers and PMCs degenerate	Hord et al., 2006
	SERK1 SERK2	*Somatic Embryogenesis Receptor-Like Kinase1 and 2*	LRR-RLKs function redundantly in sporophytic development controlling male gametophyte production and specify somatic cell identify	*serk1* serk2 double mutants are male sterile and lack tapetal layer; single mutants show no phenotype	Albrecht et al., 2005 Colcombet et al., 2005

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	ER *ERL1* *ERL2*	*ERECTA receptor-like kinase*	ER-family of LRR-RLKs act redundantly in cell differentiation during early anther development and tapetal development	Mutants produce viable pollen, however anthers do not dehisce and pollen grains adhere to anther walls	Hord et al., 2008
	MPK3 *MPK6*	*Mitogen-activated protein kinase (MAPK)*	MAPK involved in cell differentiation during anther development; acts redundantly to ER-family	Disorganized tapetal layer, reduced fertility	Hord et al., 2008
	RPK2	*Receptor-Like protein Kinase2*	Regulates tapetal development and degradation	Male sterility; limited pollen grains are clustered together, defective dehiscence	Mizuno et al., 2007
	SGC	*SGC Lectin RLK*	Perceives signals in oligosaccharide signal during early pollen development ensuring continuation of pollen development	*sgc* mutant is male sterile, collapsed pollen grains 'glued' together	Wan et al., 2008
Arabidopsis	Tapetal development				
	DYT1	*DYSFUNCTIONAL TAPETUM*	basic helix-loop-helix (bHLH) TF involved in tapetal gene regulation, regulates *AMS* and *MS1*	Male-sterile *dyt1* mutant exhibits a highly vacuolated tapetum and cytokinesis rarely occurs	Zhang et al., 2006

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	MYB33 *MYB65*	*GAMYB-like*	Involved in gibberellin (GA) signaling and subject to post transcriptional regulation by microRNAs, acts redundantly in anther development	Double mutant *myb33* myb65 is conditional male sterile; high light or low temperatures increase fertility and pollen development is limited under poor growth conditions	Millar and Gubler, 2005
	AMS	*ABORTED MICROSPORE*	bHLH TF involved in tapetal development and postmeiotic microspore formation; binds to 6 base pair consensus motif with gene targets involved in transportation of lipids, oligopeptides, and ions, fatty acid synthesis, metabolism, flavonal accumulation, substrate oxidation, methyl-modification, and pectin dynamics	Male-sterile *ams* mutant due to premature tapetal and microspore degeneration	Sorensen et al., 2003 Xu et al., 2010
	ASHR3	Arabidopsis SET-domain protein	Arabidopsis SET-domain protein; *AMS* targets *ASHR3* to chromatin regulating genes involved in stamen development	Overexpression of *ASHR3* causes growth arrest causing degenerated anthers leading to male sterility	Thorstensen et al., 2008
	MS1	*MALE STERILITY1*	Plant Homeodomain (PHD) TF; regulates late tapetal gene expression and pollen wall deposition	*ms1* mutants have variable pollen phenotypes ranging from no viable pollen; altered secretion from tapetum and program cell death (PCD)	Wilson et al., 2001 Ito et al., 2007 Yang et al., 2007

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	MYB99	*MYB99* transcription factor	Targeted by *MS1*, expressed after tetrad stage	T-DNA insertional mutant for MYB99 had reduced fertility and thinner tapetum compared to wild-type	Alves-Ferreira et al., 2007
Arabidopsis	Pollen wall formation and Exine Formation and Patterning				
	CalS5	*CALLOSE SYNTHASE5*	Callose synthase necessary for exine formation, not necessary for pollen viability	*cals5* mutants have reduced fertility because of absent callose deposition during microsporogenesis	Dong et al., 2005
	DTF1	*Tapetal development and function1*	Encodes putative *R2R3 MYB* TF controlling callose dissolution and plays a role in tapetal differentiation	Male-sterile *dtf1* mutant results in complete male sterility because of disfunction and irregular division of tapetum	Zhu et al., 2008
	A6	β-1,3-glucanase	Anther specific glucanase may be a part of callase enzyme complex		Hird et al., 1993
	MYB103	*R2R3 MYB TF*	Regulates tapetal development	Insertional mutant results in complete male sterility, tapetum becomes highly vacuolated and prematurely degenerates; Male-sterile mutant *ms188* is an allele of *Atmyb103*, defective tapetal development and callose deposition	Li et al., 2007 Zhang et al., 2007
	TDE1	*TRANSIENT DEFECTIVE EXINE 1* (*DE-ETIOLATED2*)	Brassinosteroid (BR) biosynthesis may regulate exine formation	Male-sterile mutant *tde1* primexine deposition effected, but normal exine forms	Ariizumi et al., 2008

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
Arabidopsis	*MS2*	*MALE STERILITY2*	Fatty acyl-Coenzyme A Reductase, involved in sporopollenin (fatty acid) biosynthesis	*ms2* mutant abnormalities observed when microspores are released from tetrads, some pollen is produced; moderate repression by cold stress	Aarts et al., 1997 Tang et al., 2011
	AtUTr1 *AtUTr3*	*UDP-glucose transporter*	Nucleotide sugar transporters of UDP-glucose necessary for late pollen development	Insertional mutants exhibited heterogeneous pollen grains ranging from normal, wild-type like pollen grains to flattened pollen grains devoid of cytoplasm	Reyes et al., 2010
	DEX1	*DEFECTIVE IN EXINE PATTERNING*	Novel plant protein localized to the plasma membrane involved in pollen wall patterning	*dex1* mutants anchoring of sporopollenin and exine patterning fails	Paxson-Sowders et al., 2001
	Tapetal programmed cell death				
	RPG1	*RUPTURED POLLEN GRAIN1*	MtN3/saliva family plasma membrane protein required for exine patterning	*rpg1* mutants result in partial fertility; abnormal deposition of sporopollenin from tetrad stage aborts microspores	Guan et al., 2008
	QRT1 *QRT2* *QRT3*	*QUARTET1* thru *3*	Required for tetrad pollen formation and degradation of PMC wall as microspores are released from tetrad, *QRT2* is involved in anther dehiscence	*qrt1* and *qrt2* mutants maintain pectin in PMC wall inhibiting microspore release from tetrad; Overexpression of At3g07970 (*QRT2*) reduced growth and male-sterility; Proposed model for microspore separation: pectin is demethylated by *QRT1* then degraded by *QRT2* and *QRT3*, loss-of-function *QRT2* or *QRT3* prevents microspore separation	Rhee et al., 2003 Rhee and Somerville, 1998 Ogawa et al., 2009

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	FLP1	*FACELESS POLLEN-1*	Lipid transfer protein; role in wax biosynthesis	*flp-1* conditional male-sterile mutant can be rescued under high humidity displayed a smooth pollen surface	Ariizumi et al., 2003
	NEF1	*NO EXINE FORMATION1*	Lipid biosynthesis	Male-sterile mutant *nef1* microspore abortion after release resulting in no pollen grains at anthesis	Ariizumi et al., 2004
	KNS1-12	*KAONASHI 1* thru *12*	Exine or callose synthesis	*kaonashi* mutants produce viable pollen show normal fertility	Suzuki et al., 2008
Arabidopsis	Hormonal regulation of pollen development and dehiscence				
	YUC1 YUC2 YUC4 YUC6	*YUC flavin monooqygenases*	Auxin biosynthesis; redundantly regulated growth, floral patterning, and indeterminacy	Overexpressed YUC gene results in auxin overproduction; disruption of single gene no obvious effect	Cheng et al., 2006 Sakata et al., 2010
	COI1	F-box protein	JA signaling in SKP1-Cullin-F-box (SCF) ubiquitin-ligase complex	*coi1-1* male-sterile mutants produce normal tetrads and microspores, but fail to express JA-regulated genes; temperature-sensitive	Devoto et al., 2002
	OPR3/ DDE1	*12-OXOPHYTODIENOIC ACID REDUCTASE/ DELAYED DEHISCENCE1*	JA biosynthesis; regulating anther development and dehiscence	*opr3* male-sterile mutant had low pollen germination, fertility rescued by JA application	Stintzi and Browse, 2000 Sanders et al., 2000
	MYB21		JA biosynthesis; regulating anther development and dehiscence	*myb21-1* null mutant male sterile	Mandaokar and Browse, 2009

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	MYB24 *MYB108*		Overlapping function in JA biosynthesis, downstream of *MYB21*; regulating anther development and dehiscence	*myb108* mutants have reduced male fertility, decreased filament length, and delayed dehiscence; *myb24* no effect of fertility	Mandaokar and Browse, 2009
	MYB32 *MYB4*	*R2R3 MYB TF*	Involved in pollen wall development	Insertional mutants *myb32* and *myb4* resulted in irregular, collapsed pollen grains	Preston et al., 2004
	ADPG1 *ADPG2*	*ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE*	Required for silique dehiscence, anther dehiscence, and floral abscission; regulated by JA, ethylene, and abscisic acid (ABA)	Triple mutant *adpg1 adpg2 qrt2* had delayed anther dehiscence in early flowers, delayed pollen release and pollination did not prevent seed set; *adpg2 qtr2*double mutant had greater delayed floral abscission than single mutants (*adpg1* no consistent role in floral abscission)	Ogawa et al., 2009
	ARF6 ARF8	*Auxin Response TF*	Regulates stamen development, anther dehiscence, and flower maturation	Double mutants are male and female-infertile, single mutants have delayed anther dehiscence	Nagpal et al., 2005
	TIR1	*Transport Inhibitor Response1*	Auxin receptor-encoding gene regulates anther dehiscence, filament elongation, pollen maturation, and contributes to auxin transport required for filament elongation	*tir1 abf* triple and quadruple mutants release pollen before completion of filament elongation; 30 to 60% pollen grains formed pollen tubes	Cecchetti et al., 2008
	AFB1 AFB2 *AFB3*	Auxin signaling F-box protein	Involved in late stamen development	*tir1 abf* triple and quadruple mutants release precociously developed pollen grains before filament elongation	Cecchetti et al., 2008

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	BES1	*BR-regulated TF*	Control multiple steps of anther, PMC, microspore, tapetal, and pollen coat development through binding to promoter regions of *SPL/NZZ*, *TDF1*, *MS2*, *MYB103*, *MS1*, and a differentially expressed target gene (*At3g23770)* of *MS1*	BR mutants male sterile, reduced pollen number, abnormal exine patterning anthers held pollen in locules, and abnormal tapetal development	Ye et al., 2010
Arabidopsis	Pollen tube growth				
	TMS1	*THERMOSENSITIVE MALE STERILE1*	Heat shock protein provides thermotolerance of pollen tubes	Knock out mutation in *TMS1* reduces pollen tube growth at 30 °C, significantly reducing fertility	Yang et al., 2009
	ACA9	*Autoinhibited Ca2+ATPases*	Plays a role in pollen tube growth and fertilization	*aca9* mutant pollen had reduced growth and high frequency of aborted fertilization, greater than 80% reduction in seed-set	Schiøtt et al., 2004
Rice	Archesporial specification and establishment of anther cell layers				
	OsMADS3	*C-Class MADS box TF*	Partial functional ortholog of Arabidopsis *AGAMOUS*, stamen specification	Loss-of-function mutants show severe defects in stamen identity	Yamaguchi et al., 2006
	OsMADS58	*C-Class MADS box* TF	Partial functional ortholog of Arabidopsis *AGAMOUS*, floral meristem determinacy and carpel morphology	RNAi mutants have severe defects if floral meristem determinacy	Yamaguchi et al., 2006
	MSP1	*MULTIPLE SPOROCYTE1*	LRR receptor-like kinase, ortholog of Arabidopsis *EMS1/EXS*	*msp1* mutants complete male sterility due to excessive male and female sporocytes, disrupted anther wall layers, and missing tapetum	Nonomura et al., 2003

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
Rice	Tapetal development				
	GAMYB	*GAMYB* TF	Positive regulator in GA signaling, essential pollen development	Loss-of-function by Tos17 retrotransposon in exon of OxGAMYB development shrunken anthers devoid of pollen and were male-sterile, female-fertile, more abnormal flowers developed at high than low temperatures	Kaneko et al., 2004
Rice	Tapetal programmed cell death				
	OsTDR	*TAPETAL DEGENERATION RETARDATION*	bHLH protein involved in tapetal development serves as trigger for programmed cell death (PCD); similar to Arabidopsis *AMS* and related to *DYT1* protein	Male-sterile mutant *tdr* delayed tapetal breakdown and middle layer collapse of microspores, not inhibited by increased temperatures	Li et al., 2006 Endo et al., 2009
	OsCP1	*Cysteine protease*	Tapetal PCD induction and pollen development, target of *TDR*	Loss-of-function mutation in *OsCP1* results in collapse of microspore after release from tetrad; T-DNA *oscp1* mutants showed defective pollen development with anthers devoid of viable pollen	Li et al., 2006; Lee et al., 2004
	OsC6	Protease inhibitor	Tapetal PCD induction, target of *TDR* (monocots)	*tdr* mutant had no expression of *Osc6* in anthers; Inhibited by increased temperatures	Li et al., 2006; Zhang et al., 2010
	OsTDL1A	*TAPETUM DETERMINANT1-like*	Co-expressed with *MSP1* in anthers and binds to *MSP1* to limit sporocyte numbers; homolog to *AtTPD1*	OsTDL1A-RNAi transformants normal anther size and fertility, however ovule had multiple megaspore mother cells (MeMC)	Zhao et al., 2008

Table 1. (continued)

Species	Gene	Long gene name	Putative function annotation	Mutant phenotype	Reference(s)
	OsUDT1	*UNDEVELOPED TAPETUM*	bHLH protein critical to tapetum development; likely ortholog of Arabidopsis *DYT1*	Loss-of-function by Tos17 retrotransposon or T-DNA in *Udt1* caused male sterility	Jung et al., 2005 Wilson and Zhang, 2009
Rice	Pollen wall formation and Exine Formation and Patterning				
	OsRAFTIN		Essential for maturation phase of pollen development (limited to cereals, no obvious dicot homolog)	*osRAFTIN1*-RNAi resulted male sterility, microspore collapse and inhibited tapetal degeneration; not inhibited by increased temperatures	Wang et al., 2003 Endo et al., 2009 Tang et al., 2011
	Os08g 0131100	*CYP703*	Plant-specific cytochrome P450 involved in lipid metabolism and direct target of *GAMYB*	Repressed in tapetum during high temperature	Endo et al., 2009
	Os02g 011000	Lipolytic enzyme	GDSL family protein involved in pollen coat formation and lipid metabolism	Repressed in tapetum during high temperature	Endo et al., 2009
	UDPase	*UDP-glucuronic acid decarboxylase*	Enzyme required to catalyze Mannose-1-P to UDP-xylose in polysaccharide synthesis; involved in conversion and transport of sugars for cell wall biosynthesis	Differentially present in fertile and male-sterile temperatures in thermosensitive genic male-sterile (TGMS) Zhu-1S	Xiao et al., 2009
	UGPase	*UDP-glucose pyrophosphorylase*	PCM meiosis and microspore development	Silencing *Ugp1* results in degeneration of PCM at early meiosis	Chen et al., 2007, 2009 Woo et al., 2008
	CCoAOMT	*Caffeoyl-coenzyme A O-methyltransferase*	Alternative pathway for ligin biosynthesis; involved in exine formation during pollen maturation	TGMS rice Zhu-1S, abundant in fertile condition	Xiao et al., 2009

CHAPTER 2: DAY TEMPERATURE INFLUENCES THE MALE-STERILITY LOCUS *ms9* IN SOYBEAN

A paper submitted to Crop Science

Christiana E. Wiebbecke[1,2], Michelle A. Graham[2,3], Silvia R. Cianzio[2], and Reid G. Palmer[2,3]

ABSTRACT

Cultivated soybean [*Glycine max* (L.) Merrill] is a major source of oil and protein in global food production. As demand for soy-based products increases, it is essential for breeders to find ways to increase yield beyond conventional methods. Alternative methods of cultivar development and improvement must be identified. Our objectives were to determine if the *ms9* locus conferring male-sterility in soybean was environmentally sensitive, and if it were, to determine which environments could be used in hybrid seed production. This was accomplished through characterization of pollen phenotypes and selfed seed-set among progeny of soybean genotype T359H (Genetic Type Collection number T359H), a heterozygous maintainer line for the *ms9* locus. T359H was evaluated in seven environmental regimens in growth chamber experiments. We used a split-plot design to test the effect of temperature regimens and male-sterile phenotypic classifications within chambers. As day temperature increased from 30 °C to 35 °C, selfed seed-set on male-sterile, female-fertile plants decreased. Conversely, night temperature affected neither fertile nor male-sterile, female-fertile selfed seed-set. This suggests that *ms9* can be used in hybrid seed production in controlled environments where day temperature is 35 °C during the

[1] Corresponding author; Monsanto Co., 800 North Lindbergh Boulevard, St. Louis, MO 63167

[2] Dep. of Agronomy, Iowa State Univ., Ames, IA 50011

[3] USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011. Joint collaboration of Monsanto, the Iowa Agric. and Home Econ. Exp. Stn., Ames, Iowa, Project No. 4403, and the USDA-ARS, Corn Insects and Crop Genet. Res. Unit., Ames, IA.

flowering. These temperature conditions and their effect on *ms9* may provide an alternative method to produce hybrid seed in soybean cultivar development programs.

The cultivated soybean [*Glycine max* (L.) Merrill] is one of the major oilseed crops of the world (Economic Research Service, 2010). The United States is the largest soybean producer in the world, planting 31.32 million hectares averaging 2,960 $kg \cdot ha^{-1}$ in 2010 (Agricultural Statistics Board, 2011). Soybeans are the largest source of protein feed and second largest source of vegetable oil in the world (Economic Research Service, 2010).

Cultivar development and breeding efforts have focused mainly on increasing soybean yield through selection of elite genotypes within heterogeneous families developed by manual cross-pollination (sexual hybridization). This approach has given an average rate of yield increase of 15.1 $kg \cdot ha^{-1}$, or 0.6%, per year since the 1920s (Specht and Williams, 1984). Higher yield, above and beyond conventional soybean cultivar development methods, could be achieved through the utilization of hybrids. Exploitation of hybrid vigor in naturally self-pollinated crops has provided yield increases in commercially important commodities. Hybrids have shown yield increase above inbred checks in rice, 15 to 45% (Li and Yuan, 2000; Virmani et al., 2003), pigeonpea, 14 to 64% (Saxena et al., 2010), and cotton, greater mean than mid-parent for key performance traits (McCarty et al., 2004a, 2004b; Jenkins et al., 2009). Hybrid seed production could become a feasible approach for soybean, which in some cases, hybrids yielded 10 to 20% more than the high-parent (Palmer et al., 2001).

Soybean mutant nuclear genes with Mendelian inheritance affecting male cells and organ development, are designated as *ms* genes (Horner and Palmer, 1995; Palmer et al.,

2001). The genes, typically recessive, are expressed in sporophytic tissues at different stages of development. The phenotypic expression is subdivided into three categories: 1) sporogenous, indicating normal stamen development, abnormality occurs during microgametogenesis; 2) structural, indicating the stamen is absent or abnormal (Johns et al., 1981); and 3) functional, indicating pollen is not produced or does not reach the stigma (Horner and Palmer, 1995). To date, male-sterile, female-fertile soybean lines have been used in research for population development with the hope of developing commercial hybrids (Lee et al., 1994). Male-sterile, female-fertile lines also have been used as a means to develop intermated populations in mass recurrent selection programs (Burton et al., 1990; Lewers and Palmer, 1997). The use of male-sterile, female-fertile lines with the objective of developing hybrid soybeans for commercial planting still remains a possibility.

In addition to male-sterile, female-fertile soybean genotypes, partial male-sterility also has been observed in soybean. Caviness et al. (1970) reported genotypes that showed variable expression of partial male-sterility, with 11 to 30% of normal pod set. The partially male-sterile trait was inherited as a single-recessive trait. Caviness and Fagala (1973) studied the effect of both day and night temperature on the expression of partial male-sterility. Complete sterility occurred when day temperature was 35 °C regardless of night temperature, which was either 27 °C or 21 °C, suggesting the primary influence on pod-set was daytime temperature. Male-sterile expression and pod set were influenced by temperature during reproductive development stages R1 through R8 as defined by Fehr and Caviness (1977).

Stelly and Palmer (1980) observed that male-sterility of the *msp* mutant was more pronounced as day and night temperatures decreased. Carlson and Williams (1985) showed that the partially male-sterile (*msp msp*) plants had increased male-fertility as night

temperature decreased. In male-sterile, female-fertile genotypes, environmental effects of male-sterile expression have also been reported. The *ms8* locus, in its homozygous recessive state (*ms8 ms8*), was an environmentally sensitive soybean male-sterility mutant, which showed an increase in male-sterility at high day and low night temperatures (Perez-Sackett, personal communication, 2011). These reports of environmental influences on expression of male-sterility suggest that each male-sterility locus may have its own optimum temperature regimen, which will need to be determined in order to be used in hybrid cultivar development.

In preliminary research conducted in Hawaii, Puerto Rico, and Iowa, we found that environmental conditions altered phenotypic expression of *ms9 ms9,* male-sterile, female-fertile mutant plants (Wiebbecke et al., unpublished data). At each of the two tropical locations, it was difficult to identify male-sterile, female-fertile plants in the field by visual inspection of the anther and pollen shedding at anthesis. These environmental effects precluded the use of *ms9 ms9* to produce hybrid seed in the alternate locations, and prompted the research reported herein. Our objectives were therefore to, characterize the phenotypic variation of pollen phenotypes and seed set among selfed progeny of soybean Genetic Type Collection number T359H, a heterozygous (H) maintainer line for the *ms9* locus in which *ms9 ms9* plants are male-sterile, female-fertile (Palmer, 2000). This would allow us to determine if the *ms9* locus conferring male-sterility in soybean was environmentally sensitive, and if it were, to determine which environments could be used for hybrid seed production. The interactions between the *ms9 ms9* genotype and different environments that could influence male-fertility and sterility expression were investigated.

MATERIALS AND METHODS

Plant Material

Soybean Genetic Type Number T359H was developed from the initial cross of T325 × L67-3483 (Palmer, 2000). T325 was identified as a germinal revertant of *w4*-mutable line (Hedges and Palmer, 1992), containing an active transposable element (Xu et al., 2010). L67-3483 originated from X-ray irradiation of the cultivar 'Clark' (Kato and Palmer, 2003). T359H has been maintained by self-pollination of *Ms9 ms9* plants, expected to produce progeny with the approximate phenotypic ratio of 3 fertile plants: 1 male-sterile, female-fertile plant.

Environmental Conditions

Temperature regimen. Growth chamber settings were selected to mimic environmental conditions at Puerto Rico, Hawaii, and Iowa locations (Table 1). This enabled us to conduct visual pollen phenotyping at the R2 stage, during each different set of simulated environments. Four experiments were conducted with seven temperature regimens to test night and day temperature effects (Table 2). Experiment 4 in chamber A, served as replicate to experiment 2 in chamber B (30 °C day/23 °C night temperature), which allowed a statistical test of random effects due to chamber differences.

Photoperiod. Daily photoperiod illumination provided by incandescent and fluorescent lamps was 18 h for week one, 16 h for week two, 14 h for week three through remainder of the first experiment in both chambers. In experiments 2 through 4, we decreased daily illumination from 14 h to 13 h at week four, a stage at which approximately two sets of trifoliolates were unfolded, the V2 vegetative stage (Fehr and Caviness, 1977). We reduced the photoperiod in effort to decrease extreme internode elongation observed in

experiment 1. For experiments 2 through 4, the 13 h photoperiod was continued through full maturity (R8, Fehr and Caviness, 1977). Plants were grown until flowering ceased for fertile plants or for male-sterile, female-fertile plants seeds were of sufficient size to count the number of pods and seeds per plant to determine the selfed seed-set of T359H progeny.

Experiments

For each of the four experiments, approximately 120 seeds were germinated using the germination towel method number 8 (Burris and Fehr, 1971). During germination, containers were kept in a growth chamber with 18 h illumination and at 30 °C. One week after germination, healthy seedlings were transplanted to each of 28 pots containing a mixture of 2 soil : 1 sand : 1 peat moss. The soil was supplied by the Bruner Farm near Ames, IA. Each 8.8-L pot contained two healthy seedlings, totaling 56 plants. Each temperature regimen (Model P6W36, Controlled Environments LTD, Manitoba, Canada) consisted of 28 pots that were randomly placed in chambers. To reduce within chamber effects during the experiments, pots were moved randomly throughout the chamber until pod fill began (R4, Fehr and Caviness, 1977). Plants were fertilized at V2 stage, with Miracle-Gro® Water Soluble All Purpose Plant Food containing 28N–8P–16K at a rate of 4 g per 1 L tap water, following manufacturers recommended rate, using approximately 300 mL per pot. Within each chamber, plants were numbered to facilitate data collection.

Phenotypic Classification

Visual classification of flowering. During R2 stage, we dissected one to two fully open flowers from each plant in the growth chamber to test pollen shedding. Flowers were dissected using tweezers. The sepals and petals were removed and then the anther was gently lifted away from the remaining parts of the flower. Once the anther was removed, it was

dusted onto a person's thumbnail to test pollen release (Fehr, 1980). Initial classification of pollen phenotype depended on deposition of pollen on the surface of the thumbnail.

Microscopy. At R2 stage, we collected reproductive buds and flowers from fertile and male-sterile, female-fertile plants approximately one day before anthesis and at anthesis. Samples were numbered according to the plant they were sampled from. Samples were placed in glass vials containing 70% ethanol for storage. Samples were used to conduct cytological observations on anther and pollen development. Anthers were dissected under a dissection microscope. Late-stage anthers were squashed in an aqueous solution of I_2KI, which stained starch within the pollen grains (Jensen, 1962). Under a light microscope, we were able to differentiate between fertile versus male-sterile pollen phenotypes. Fertile pollen grains accumulated stain and were tri-colporate in shape. Pollen grains from male-sterile, female-fertile plants varied in phenotype, typically displaying a mixture of: normal fertile pollen grains, large grains typical of 2n pollen, and small unstained grains, devoid of starch.

Selfed seed-set at harvest. Plants were grown in the growth chamber until maturity for fertile plants or R4 stage for male-sterile, female-fertile plants, when seeds were of ample size in pods to be counted without shelling. Between the 11^{th} and 13^{th} weeks after transplanting, we collected pods from each plant and counted seeds within pods.

Progeny Test

Twenty-three male-sterile, female-fertile plants from chambers A and B grown during experiment 1, were moved to the USDA greenhouse, located at the Iowa State University campus, to allow seeds to fully mature (R8 stage). Plants had between two and 21 seeds, that

were planted at the Bruner Farm in May 2010. At maturity (R8 stage), we classified male-sterile phenotypes based on selfed seed-set.

Statistical Analysis

Chi square tests (Snedecor and Cochran, 1989) were conducted to assess if the observed male-sterile phenotype of *ms9* deviated from the expected 3 : 1 Mendelian segregation ratio of fertile : male-sterile for a single recessive locus. Sample variances plotted against mean seed count were inspected for each group of observations formed by combinations of male-sterile phenotype and temperature regimen, and they were found to be heterogeneous. In order to combine data for the analysis of variance (ANOVA), seed counts of individual plants were transformed using a square root transformation to stabilize variances of count data (Snedecor and Cochran, 1989).

To determine environmental factors that influenced seed production for the *ms9* locus, we analyzed the data using a split-plot experimental design in which whole-plots were temperature regimens across growth chamber and sub-plots were male-sterile phenotypic classification within chambers. PROC MIXED in SAS version 9.2 (SAS Institute Inc., 2008) was used to accommodate unbalanced samples of fertile and male-sterile, female-fertile plants across temperature regimens. Temperature regimen and male-sterile phenotype were considered fixed effects and growth chamber was considered a random effect. Because the interaction between male-sterile phenotype and temperature regimen was significant, we then used contrast statements to test significance of temperature and photoperiod effects within male-sterile phenotypes. PROC MIXED also was used to determine estimated means and standard errors for each level combination of male-sterile phenotype and temperature regimen.

RESULTS

Inheritance. Progeny of self-pollinated soybean Genetic Type Number T359H plants segregated in a Mendelian ratio of 3 fertile plants : 1 male-sterile, female-fertile plant, as expected. Plants grown in all growth chamber experiments, regardless of temperature and photoperiod regimens, fit the expected segregation ratio for a single-recessive locus (Table 3). In the progeny test of selfed seed collected from male-sterile, female-fertile plants from experiment 1 growth chambers A and B (30 °C day/21 °C night; 35 °C day/17 °C night), all progeny were male-sterile, female-fertile.

Phenotypic Screening

Anther phenotypes. By visual inspection, we recorded that male-sterile, female-fertile plants had pallid, shriveled anthers, which did not dehisce when tapped. Anthers of fertile progeny appeared light yellow, fluffy, and dehisced small clouds of pollen. With microscopy, we were able to differentiate two male-sterile anther phenotypes. In the first class, pollen remained fixed within male-sterile anther sacs, while in the second, anthers were devoid of pollen (Figure 3A and 3B). The second anther class was associated with the five male-sterile, female-fertile plants that did not generate pollen. From fertile plants, we observed engorged anthers that would allow pollen to disperse freely from the anther (Figure 3C).

Pollen phenotypes. During microscopy, male-sterile, female-fertile plants were classified by the presence (Figure 1A-B and 2A) or absence of pollen (Figure 3B). Within pollen producing plants, further classifications were made based on pollen phenotype. The majority of male-sterile, female-fertile plants produced a blend of three types of pollen: normal-viable (tri-colporate and intensely stained), 2n appearance (large, round and

brownish-black in color), and nonviable (clear, flattened) pollen grains (Figure 1A-B and 2A). These grains often formed clusters, adhering to each other (Figure 1A). Fertile plants had normal-viable pollen grains that were tri-colporate and darkly stained with I_2-KI (Figure 1C and 2B). As previously mentioned, five plants failed to produce any pollen. These plants set no seed or vestigial pods. Further, the five plants were observed randomly distributed in each of four temperature regimens (35 °C day/17 °C night; 30 °C day/12 °C night; 30 °C day/18 °C night; 35 °C day/18 °C night). These observations were consistent with the visual assessment of anther phenotypes.

Pod-fill. Phenotypic differences between male-sterile, female-fertile and fertile progeny of T359H were most apparent at maturity. Male-sterile, female-fertile plants retained their dark green-leathery leaves and developed woody stems. These plants continued to produce flowers throughout the final six weeks of the experiments and did not transition completely through the pod fill stage. This was in clear contrast to their fertile counterparts, which began pod fill, set seed, and progressed to maturity.

When male-sterile, female-fertile plants did set seed, pods were malformed, often had ovule abortions, and the few seeds produced were larger than seeds produced on fertile plants. Pods on male-sterile, female-fertile plants were typically located at the lower section of the main stem, and on branches of plants, with few pods distributed on the main stem. In addition, the majority of male-sterile, female-fertile plants set vestigial pods that were scattered throughout the length of the main stem. In seven of the eight growth chamber experiments (excluding experiment 2 chamber B, 30 °C day/23 °C night), a total of 24 male-sterile, female-fertile plants set no seed. Of these, 19 plants set vestigial pods and five did

not set pods. Fertile plants did not set vestigial pods and set a greater number of two-, three-, and four-seeded pods distributed on the main stem and branches.

Effects of temperature on selfed seed-set. The most dramatic decrease in selfed seed-set on male-sterile, female-fertile plants occurred when day temperature was increased from 30 °C to 35 °C (Figure 4, $P = 0.0331$). The greatest selfed seed-set on male-sterile, female-fertile plants occurred at 30 °C day temperature, with 16 seeds per plant on average compared to an average of 1 seed per plant when day temperature was 35 °C. Statistical analysis detected a difference between chambers when temperature regimen and photoperiod were held constant ($P = 0.0138$). We determined the chamber effect that was detected during statistical analysis was attributed to additional floral sampling for microscopy during experiment 4. Extra sampling requirements decreased flowers available for self-pollination and decreased average seed set in experiment 4 chamber A compared to the previous replication of that temperature regimen, experiment 2 chamber B. This rationalization provided justification to pool data for the 30 °C day/23 °C night temperature regimen across chambers (D. Nordman, personal communication, 2011).

There was no effect on selfed seed-set adjusting day temperature of fertile progeny ($P = 0.657$) (Figure 4). Changing night temperature, while holding day temperature constant at 30°C, neither affected selfed seed-set of male-sterile, female-fertile ($P = 0.954$) nor fertile plants ($P = 0.195$). For experiments 2 through 4, we shortened daily photoperiod, from 14 h to 13 h, which had no effect on selfed seed-set of male-sterile, female-fertile nor fertile plants ($P = 0.266$).

DISCUSSION

Inheritance. Soybean Genetic Type Number T359H was formed by hybridization of T325 × L67-3483 (Palmer, 2000). Because of genetic instability in parental genotypes of T359H (active transposable element and X-ray irradiation), we speculated that the differential expression of male-sterile phenotype we observed in our preliminary research may have been the result of reversion of male-sterile to male-fertile. To test this theory, we conducted a progeny test, which demonstrated all progeny produced by self-pollination of male-sterile, female-fertile plants gave plants with a male-sterile phenotype. This suggests reversion caused by further genomic changes is unlikely, further supporting our hypothesis that environmental factors influence the expression of the male-sterile phenotype for the *ms9* locus in soybean.

Thermosensitive Male-Sterility Loci

There was no effect of night temperature on expression of the male-sterile phenotype of *ms9*. This result is contrary to other environmental influence studies utilizing the *msp* locus where night temperature did influence male-sterility expression (Carlson and Williams, 1985). Unpublished results of the *ms8* locus indicated both day/night temperature had effects on the phenotype of the plants (Perez-Sackett, personal communication, 2011). Lack of night temperature effects on *ms9* seed set and pollen phenotype is similar to results with the partial male-sterile mutant described by Caviness and Fagala (1973). Male-sterility conferred by *ms9* locus was maintained at high (35 °C) day temperature, where as male-sterile, female-fertile plants would revert to partial fertility at low day temperature (30 °C). Expression of male-sterility at the *ms9* locus is comparable to thermosensitive genic male-sterility identified in rice, in which hybrid seed producers used different environments for male-

sterile line maintenance and hybrid seed production while controlling temperature (Li and Yuan, 2000; Virmani et al., 2003).

Phenotypic expression of male-sterility at the *ms9* locus corresponds to all three categories described by Horner and Palmer (1995). Male-sterile, female-fertile T359H progeny had stamen and anthers with normal appearance, but were defective pollen production, indicating sporogeneous phenotypic expression (Figure 3A). Five plants distributed randomly among four of seven temperature regimes set no seed and displayed a unique no pollen, pale-anther phenotype indicating structural phenotypic expression of male-sterility (Figure 3B). Normal-viable pollen was not produced consistently, showing functional phenotypic expression (Figure 1A and 1B). The diverse phenotypic expression of male-sterility at the *ms9* locus indicates breakdown of pollen development during the later stages of microgametogenesis. Late-stage failure allows some functional pollen to complete microgametogenesis and enables male-sterile line maintenance via self-pollination in controlled environments.

The observation and confirmation of thermosensitivity at *ms9* male-sterility locus in soybean suggests that this male-sterility locus could be used in hybrid cultivar development similar to methods utilized in hybrid rice production (Li and Yuan, 2000; Virmani et al., 2003). Identifying the critical sterile-inducing temperature for the *ms9* locus could enable soybean seed producers to grow male-sterile maintainer lines. Maintenance of male-sterile lines could be achieved by growing male-sterile, female-fertile plants in controlled 30 °C day temperature environments, while eliminating insect pollinators to reduce potential outcrossing. This novel approach for soybean male-sterile line maintenance provides a system to generate uniform lines for use as the female parent in hybrid cultivar development.

Soybean breeders could then utilize hybrid seed production by growing male-sterile, female-fertile plants in controlled 35 °C day temperature environments with a complementary male line while introducing insect pollinators to facilitate cross pollination. Critical timing of day temperature influencing *ms9* partial fertility reversion remains to be determined and would optimize this system for use in soybean hybrid cultivar development. Ultimately, protocols could be devised for male-sterile line maintenance and for hybrid seed production, which could overcome some key challenges when implementing soybean hybrid production. These methods along with the detection of heterotic associations for soybeans may lead to the broader use of hybrid soybean cultivar development and further enhance genetic gain.

Acknowledgements

This project was a joint contribution from Monsanto Co., the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 4403, and from the USDA Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, and supported by the Hatch Act and the State of Iowa. The mention of a trademark or proprietary product by Monsanto Co., Iowa State University, or the USDA, and the use of the name by Monsanto Co., Iowa State University, or the USDA does not imply its approval to the exclusion of other products that may be suitable. We would like to express our appreciation to Dan Nordman for his assistance with statistical design and analysis. Additionally, we would like to acknowledge Allison Pappas and Randall Den Adel for their assistance capturing microphotograph images.

REFERENCES

Agricultural Statistics Board. 2011. Crop Production. National Agriculture Statistics Service (NASS), USDA. February 2011.

Burris, J.S., and W.R. Fehr. 1971. Methods for evaluation of soybean hypocotyl length. Crop Sci. 11:116-117.

Burton, J.W., E.M.K. Koinange, and C.A. Brim. 1990. Recurrent selfed progeny selection for yield in soybean using genetic male sterility. Crop Sci. 21:31-34.

Carlson, D.R., and C.B. Williams, III. 1985. Effect of temperature on the expression of male sterility in partially male-sterile soybean. Crop Sci. 25:646-648.

Caviness, C.E., and B.L. Fagala. 1973. Influence of temperature on a partially male-sterile soybean strain. Crop Sci. 13:503-504.

Caviness, C.E., H.J. Walters, and D.L. Johnson. 1970. A partially male-sterile strain of soybean. Crop Sci. 10:107-108.

Economic Research Service. 2010. Soybean and oil crops briefing room: Overview. < http://www.ers.usda.gov/Briefing/SoybeansOilCrops/>. (accessed on27Feb2011; verified 04May2011). USDA, ERS.

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa Agri. Home Econ. Expt. Stn., Iowa State Univ., Ames, IA.

Fehr, W.R. 1980. Soybean, p. 589-599. *In* W.R. Fehr and H.H. Hadley. (ed.) Hybridization of crop plants. Am. Soc. of Agron., Inc., Madison, WI.

Hedges, B.R., and R.G. Palmer. 1992. Inheritance of malate dehydrogenase nulls in soybean. Biochem. Genet. 30:491-502.

Horner, H.T., and R.G. Palmer. 1995. Mechanisms of genic male sterility. Crop Sci. 35:1527-1535.

Jenkins, J.N., J.C. McCarty, J. Wu, and O.A. Gutierrez. 2009. Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. Euphytica 167:397-408.

Jensen, W.A. 1962. Botanical histochemistry. Freeman, San Fransisco, CA.

Johns, C.W., X. Delannay, and R.G. Palmer. 1981. Structural sterility controlled by nuclear mutations in angiosperms. The Nucleus 24:97-105.

Kato, K.K., and R.G. Palmer. 2003. Genetic identification of a female partial-sterile mutant in soybean. Genome 46:128-134.

Lee, D.L., C.A. Caba, J.E. Specht, and G.L. Graef. 1994. Analysis of cytoplasmic diversity in an outcrossing population of soybean. Crop Sci. 34:46-50.

Li, J., and L. Yuan. 2000. Hybrid rice: Genetics, breeding, and seed production. Plant Breed. Rev. 17:15-158.

Lewers, K.S., and R.G. Palmer. 1997. Recurrent selection in soybean. Plant Breed. Rev. 15:275-313.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004a. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. Crop Sci. 44:1226-1230.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004b. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: II. Genetic effects and genotypic values. Crop Sci. 44:1231-1235.

Palmer, R.G. 2000. Genetics of four male-sterile, female-fertile soybean mutants. Crop Sci. 40:78-83.

Palmer, R.G., J. Gai, H. Sun, and J.W. Burton. 2001. Production and evaluation of hybrid soybean. Plant Breed. Rev. 21:263-307.

SAS Institute Inc. 2008. SAS/STAT 9.2 user's guide. Cary, NC.

Saxena, K.B., R. Sultana, N. Mallikarjuna, R.K. Saxena, R.V. Kumar, S.L. Sawargaonkar, and R.K. Varshney. 2010. Male-sterility systems in pigeonpea and their role in enhancing yield. Plant Breed. 129:125-134.

Snedecor, G.W., and W.G. Cochran. 1989. Statistical methods. 8th ed. Iowa State Univ. Press, Ames, IA.

Specht, J.E., and J.H. Williams. 1984. Contribution of genetic technology to soybean productivity – retrospect and prospect, p.49-74. *In* W.R. Fehr (ed.) Genetic contributions to yield gains of five major crop plants. Am. Soc. of Agron., Inc., Madison, WI.

Stelly, D.M., and R.G. Palmer. 1980. A partially male-sterile mutant line of soybeans, *Glycine max* (L.) Merr.: Characterization of the *msp* phenotype variation. Euphytica 29:539-546.

Virmani, S.S., Z.M. Sun, T.M. Mou, A.J. Ali, and C.X. Mao. 2003. Two-line hybrid rice seed production, p. 53-64. *In* B. Hardy (ed.) Two-line hybrid rice breeding manual. International Rice Research Institute. Los Baños, Philippines.

Xu, M., H.K. Brar, S. Grosic, R.G. Palmer, and M.K. Bhattacharyya. 2010. Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean, [*Glycine max* (L.) Merr.]. Genetics 184:53-63.

Table 1. Temperature data collected by weather stations at Monsanto MSP research farms near Kunia, HI, and Isabela, PR. Huxley, IA temperature data were obtained from the National Climate Data Center for Ames, IA. Minimum, maximum, and average daily temperatures were reported for the months when T359H was grown at that location. The R2 reproductive stage (Fehr and Caviness, 1977) represents the approximate time when plants were in full flower.

			Daily Temperature			
Location	**Year**	**Month**	**Minimum (°C)**	**Maximum (°C)**	**Average (°C)**	**R2 Stage†**
Kunia, HI	2008	Sept. 25th on	19.6	31.9	25.8	
Kunia, HI	2008	Oct	19.6	32.0	25.4	Late Oct.‡
Kunia, HI	2008	Nov	18.4	29.7	24.0	
Kunia, HI	2008	Dec	18.0	29.7	23.1	Mid Dec. §
Kunia, HI	2009	thru Jan. 30th	13.4	27.1	21.4	
Isabela, PR	2009	Apr. 29th on	21.0	27.5	N/A¶	
Isabela, PR	2009	May	21.3	29.4	24.2	
Isabela, PR	2009	June	22.1	30.3	25.3	Early June
Isabela, PR	2009	July	22.5	31.1	26.5	
Isabela, PR	2009	thru Aug. 27th	22.5	30.3	26.3	
Huxley, IA	2009	May	1.6	30.0	15.5	
Huxley, IA	2009	June	8.3	34.4	21.0	
Huxley, IA	2009	July	10.0	30.5	20.0	Early July
Huxley, IA	2009	Aug	7.2	33.3	20.5	

† R2 Stage, Fehr and Caviness, (1977)

‡ Late September Kunia, HI, planting

§ Late October Kunia, HI, planting

¶ Not applicable

Table 2. PROC MIXED in SAS version 9.2 (SAS Institute Inc., 2008) was used to estimate mean seed count and standard error of male-sterile, female-fertile (ms) and fertile (F) phenotypes among T359H progeny across the eight growth chamber experiments. Contrast statements on square root transformed data were used to test significance of photoperiod, day, and night temperature effects within male-sterile phenotypes. Each growth chamber contained 28 pots with two plants per pot, a total of 56 plants.

Expt.	Growth chamber	Day (°C)	Night (°C)	Photoperiod†	Male-sterile phenotype	No. plants	Mean seed count/ plant	SE
2	A	30	12	13	F	44	76	6.40
2	A	30	12	13	ms	12	16	7.91
3	A	30	15	13	F	41	58	6.44
3	A	30	15	13	ms	15	17	7.53
1	B	35	17	14	F	43	43	6.41
1	B	35	17	14	ms	13	2	7.77
3	B	30	18	13	F	41	41	6.44
3	B	30	18	13	ms	41	20	7.53
4	B	35	18	13	F	15	53	6.44
4	B	35	18	13	ms	15	1	7.53
1	A	30	21	14	F	37	45	6.52
1	A	30	21	14	ms	19	11	7.18
2 & 4‡	B & A	30	23	13	F	88	46	4.53
2 & 4‡	B & A	30	23	13	ms	24	18	5.62

† Photoperiod 14 = 18 h day/6 h night (1 wk); 16 h day/8 h night (2 wk); 14 h day/10 h night (to end of experiment): Photoperiod 13 = 18 h day/6 h night (1 wk); 16 h day/8 h night (2 wk); 14 h day /10 h night (3 wk); 13 h day/11 h night (to end of experiment).

‡ 30 °C day/23 °C night replicated in experiment 2 chamber B and experiment 4 chamber A were pooled. However, there was a chamber effect ($P = 0.0138$) due to increased floral sampling during experiment 4.

Table 3. Segregation ratios of *ms9* locus in soybean T359H progeny comparing fertile versus male-sterile, female-fertile phenotype while varying day and night temperature combinations. Neither day nor night temperature affected Mendelian segregation of the single-recessive male-sterility *ms9* locus in soybean.

			Number of plants			
Expt.	**Growth chamber**	**Day/night (°C)**	**Fertile**	**Sterile**	**χ2 (3:1)**	***P***
1	A	30/21	37	19	2.38	0.123
1	B	35/17	43	13	0.10	0.752
2	A	30/12	44	12	0.38	0.538
2	B	30/23	48	8	3.43	0.064
3	A	30/15	41	15	0.10	0.752
3	B	30/18	41	15	0.10	0.752
4	A	30/23	40	16	0.38	0.538
4	B	35/18	41	15	0.10	0.752

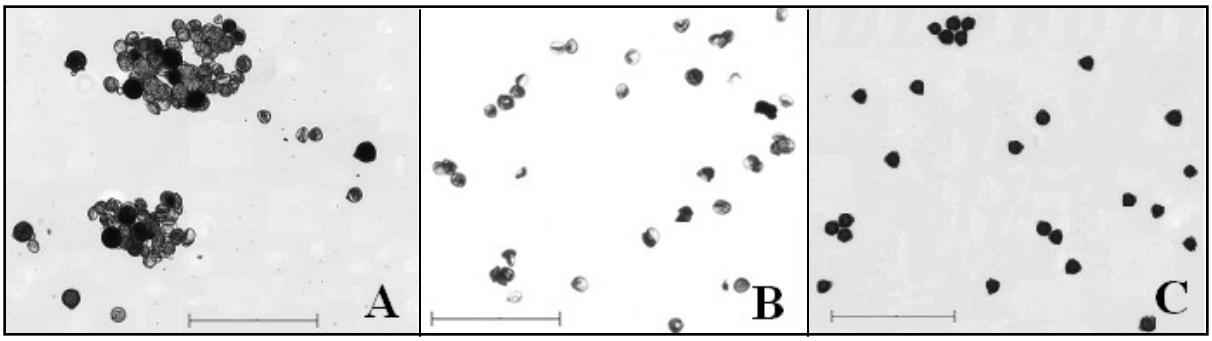

Figure 1. Photomicrographs of soybean pollen grains in I_2-KI solution were used to classify normal or abnormal pollen phenotypes. Pollen samples were collected from the selfed progeny of soybean Genetic Type Number T359H, the heterozygous (H) maintainer line for the male-sterility *ms9* locus. Scale bars equal 200 μm. (A) Pollen grains from male-sterile, female-fertile plant no. 4-A-1. (B) Pollen grains from male-sterile, female-fertile plant no. 4-B-56. (C) Pollen grains from fertile plant no. 4-B-20.

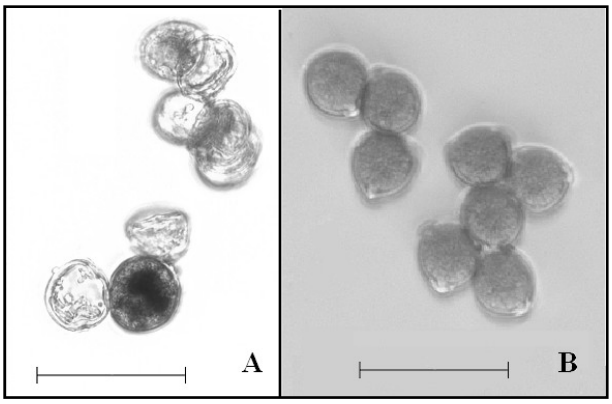

Figure 2. Photomicrographs of soybean pollen grains in I_2-KI solution were used to classify normal or abnormal pollen phenotypes. Pollen samples were collected from selfed progeny of T359H, heterozygous for *ms9* locus. Scale bars equal 50 μm. (A) Pollen grains from male-sterile, female-fertile plant no. 1-B-39. (B) Pollen grains from fertile plant no. 4-B-7.

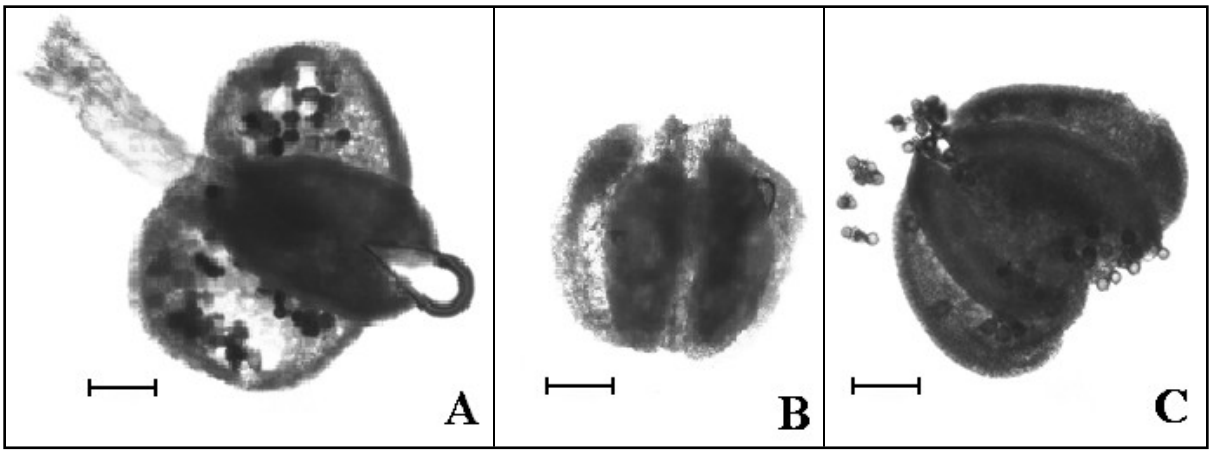

Figure 3. Photomicrographs of squashed soybean anthers in I_2-KI solution classified as normal or abnormal anther phenotypes. Anther samples were collected from selfed progeny of T359H, heterozygous for *ms9* locus. Scale bars equal 100 µm. (A) Squash of mature male-sterile anther showed that most of the pollen remained in the anther sacs, plant no. 4-A-13. (B) Squash of mature male-sterile anther void of pollen grains, plant no. 4-A-22. (C) Squash of mature fertile anther showed release of engorged pollen grains, plant no. 4-A-48.

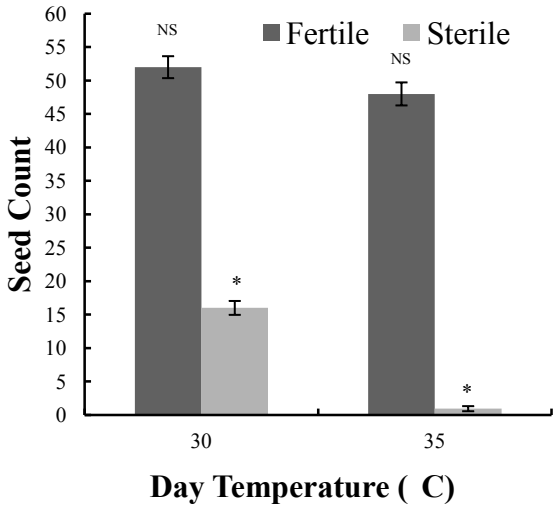

Figure 4. Mean seed counts of male-sterile, female-fertile (*ms9 ms9*) plants decreased as day temperature increased from 30 °C to 35 °C (* indicates significance, $P = 0.0331$). In contrast, mean seed count of fertile (*Ms9 Ms9* or *Ms9 ms9*) plants showed no change in mean seed count (^{NS} indicates not significant, $P = 0.657$). The error bars represent standard error of the mean calculated by PROC MIXED in SAS version 9.2 (SAS Institute Inc., 2008). Contrast statements were used to compare the effect of day temperature on seed set. Daily photoperiod was 13 h of illumination from V2 stage through maturity. Each growth chamber contained 28 pots with two plants per pot, totaling 56 plants.

CHAPTER 3: COMPUTATIONAL IDENTIFCATION OF CANDIDATE GENES IN SOYBEAN MALE-STERILITY LOCI

A paper for submission to Plant Genome

Christiana E. Wiebbecke[1], Silvia R. Cianzio[1], Reid G. Palmer[1,2], and Michelle A. Graham[2,3]

ABSTRACT

Researchers have long been interested in floral development and pollination, as successful pollination has a direct impact on maintaining and enhancing high crop yields. In particular, soybean breeders are interested in understanding phenotypic differences in male-sterility loci that could be manipulated for use in hybrid seed production. With the recent release of multiple genomic tools for soybean, we can now use bioinformatic approaches to identify candidate male-sterility genes and compare the function of multiple male-sterility loci. The objectives of this research were to use computational analyses to identify candidate genes within known male-sterility loci and determine if transcription factors play a role in conditioning environmental sensitivity across multiple loci. Our bioinformatic analysis identified 23 candidate genes in three of the four molecularly mapped environmentally sensitive male-sterility loci (*ms9*, *msp*, and *ms3*) and the environmentally stable *ms2* locus. Our computational analysis also identified known transcription factor binding sites (TFBS) shared across multiple environmentally sensitive male-sterility loci, as well as TFBS unique to each male-sterility locus. Ultimately, these bioinformatic approaches have identified candidate genes for male sterility that could be used in marker assisted selection programs.

[1] Dep. of Agronomy, Iowa State Univ., Ames, IA 50011

[2] USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011. Joint collaboration of the Iowa Agric. and Home Econ. Exp. Stn., Ames, Iowa, Project No. 4403, and the USDA-ARS, Corn Insects and Crop Genet. Res. Unit., Ames, IA.

[3] Corresponding author (Michelle.Graham@ARS.USDA.GOV)

Identifying the TFBS that regulate male-sterility genes will enable the improvement of male-sterile systems in soybean and could ultimately lead to a more broad use of hybrids in soybean cultivar development programs.

INTRODUCTION

For decades researchers have been interested in floral development because it has a direct impact on yield. A key step in seed production is successful pollination, therefore, understanding floral characteristics and the reproductive biology provides insight into maintaining or enhancing high crop yields (Wilcock and Neiland, 2002). In self-pollinated crops, the amount of out-crossed seed produced is directly related to the magnitude of attraction by insect pollinators (Palmer et al., 2009). Plant breeders must appreciate the importance of crop-pollinator relationships, such as nectar production and floral display, as it relates to potential hybrid yield and yield stability (Palmer et al., 2009). Self pollinated crops have shown heterosis for yield in rice, 15 to 45% (Li and Yuan, 2000; Virmani et al., 2003) and pigeonpea, 14 to 64% (Saxena et al., 2010). Similarly, cotton hybrids yield greater than the mid-parent for critical agronomic traits have been reported (McCarty et al., 2004a, 2004b; Jenkins et al., 2009).

In soybean, complete sterility first was observed in the progeny of Manchu soybean (*Glycine max* (L.) Merr.) (Owen, 1928). Sterile plants did not set seed, remained green, and held their leaves. Failure to complete maturity occurred because the traditional sink (seeds) was not there to accumulate the starch produced during the extended reproductive phase (Owen, 1928). Investigation by Hadley and Starnes (1964) determined sterility of soybean lines T241 and T242 were associated with homozygous recessive genes (*st2* and *st3*) causing

asynapsis during meiosis (Hadley and Starnes, 1964). In addition to the sterile phenotype described above, partial-female sterility (Pereira et al., 1997), partial-male sterility (Caviness et al., 1970; Stelley and Palmer, 1980), and male sterility (Horner and Palmer, 1995; Palmer et al., 2001) also have been observed in soybean. Since soybean is a self-pollinated crop, improving understanding of male-sterility loci has been of particular interest because of potential use in hybrid seed production. Although sterility first was discovered in soybean over 80 years ago and research to study the inheritance, phenotypic variation, cytology, and location of male-sterility loci has taken place, the molecular mechanisms explaining male sterility in soybean are still poorly understood.

Genetic and molecular studies in Arabidopsis, snapdragon, tobacco, maize, and petunia have identified complex arrays of gene interactions involved in anther morphology and pollen grain differentiation (Goldberg et al., 1993). More specifically, male-sterile mutant screens in the model plant systems of Arabidopsis and rice have identified genes involved in establishment of anther cell layers, archesporial specification, tapetal development, pollen wall formation, exine formation and patterning, programmed cell death (PCD) of the tapetum, hormonal regulation of pollen development, and anther dehiscence (Scott et al., 2004; Wilson and Zhang, 2009).

Similarly, environmentally-sensitive genic male sterility has been identified in several self-pollinated crop systems including wheat (Jan, 1974), barley (Sharma and Reinbergs, 1976), rice (Sun et al., 1989; Virmani et al., 2003), and oilseed rape (Xi et al., 1997). Researchers identified specific genes that play a role in thermosensitive genic-male sterility (TGMS). In Arabidopsis, the *THERMOSENSITIVE MALE STERILE1* (*TMS1*) gene encodes a heat shock protein that provides tolerance to high temperatures during pollen-tube growth

and development (Yang et al., 2009). *TMS1* knock out mutants have decreased male fertility. In rice, the *ms-h* gene encoding a UDP-Glucose Pyrophosphorylase (*UGPase1*, Woo et al., 2008) is essential for callose deposition during pollen mother cell (PMC) meiosis and microspore development (Chen et al., 2007, 2009). Temperature-sensitive splicing of *Ugp1* causes male-sterile plants to revert to fertility after a period of low temperatures. In wheat, *NPK1* (nucleus- and phragmoplast-localized protein kinase 1), plays a role in cytoskeletal organization and male meiotic cytokinesis (Tang et al., 2011). Tang et al. (2011) observed an increase in expression of a *NPK1* homolog in wheat anthers after cold stress and histological observations showed abnormal separation of dyads during male meiosis I. Another example of environmental sensitivity is repression of auxin biosynthesis genes during periods of increased temperatures, which causes an anther-specific reduction in auxin, thus maintaining male sterility in Arabidopsis and barley (Sakata et al., 2010). These examples of environmentally influenced male-sterility genes may offer insight into potential candidate genes for male sterility in soybean.

In soybean, five male-sterility loci have been mapped: *ms9* (Cervantes-Martinez et al., 2007), *ms3* (Cervantes-Martinez et al., 2009), *ms8* and *msp* (Frasch et al., 2010), and *ms2* (Cervantes-Martinez et al., 2007). Four of the genes (*ms9*, *ms3*, *ms8*, and *msp*) have shown environmental sensitivity. In spite of cytological differences (Palmer et al., 2001), the genes causing male sterility at these loci may be similar, since they all experience atypical floral development during microgametogenesis.

With the release of the soybean genome sequence (Glyma 1, Schmutz et al., 2010) and soybean genome browser (www.soybase.org), these regions can be visualized and potential candidate genes can be identified. In addition, the recent publication of the soybean

RNA-Seq atlas (Severin et al., 2010), provides expression data for all predicted soybean genes in a total of fourteen tissue types (root, nodule, young leaf, flower, pod, and seed including different time points during pod and seed development). The RNA-Seq atlas (Severin et al., 2010) allows researchers to evaluate expression data of a gene of interest and determine if it is expressed in specific tissues, developmental stages, or is constitutively active. This information could provide a better understanding of phenotypic differences in male sterility.

Using the resources described above, our intent is to improve the understanding of male sterility in soybean. The objectives of this study were to use computational methods to 1) identify candidate genes within male-sterility loci with homology to known male-sterility genes from other species, 2) characterize transcription factor binding sites (TFBS) common across environmentally sensitive male-sterility loci or unique TFBS to each male-sterility locus, and 3) compare TFBS between male-sterility loci and their homeologous regions to determine if changes in gene expression or TFBS lead to subfunctionalization (Force et al., 2005, Yang et al., 2006). Subfunctionalization could explain why the homeologous regions have no known roles in male sterility. The results of these bioinformatics analyses will enable future laboratory experiments to confirm the function of candidate genes in male-sterility loci.

MATERIALS AND METHODS

Detection of Genome Regions Involved in Male-Sterility

Five male-sterility loci previously have been mapped; *ms9* (Cervantes-Martinez et al., 2007), *ms3* (Cervantes-Martinez et al., 2009), *ms8* and *msp* (Frasch et al., 2010), and *ms2*

(Cervantes-Martinez et al., 2007). The male-sterile phenotype of four of these loci is influenced by the environment (*msp*, Stelley and Palmer, 1980; *ms9*, Wiebbecke et al., 2011; *ms8*, Perez-Sackett et al., 2011; *ms3*, Wiebbecke et al., unpublished data). In contrast, *ms2*, appears to have a stable male-sterile phenotype (Cervantes-Martinez et al., 2007 and Ortiz-Perez et al., 2006). The closest flanking SSR markers for each male-sterility locus were used to query the SoyBase database (http://www.soybase.org). The genes closest to the flanking markers defined the range of candidate genes for each male-sterility locus (Table 1). We then downloaded the entire genomic sequence between markers, as well as the sequence for each individual gene for each defined region (Table 1).

Cluster Analysis of Male-Sterility Regions

Computational identification of homeologous genes was performed by comparing the 1,198 predicted gene sequences from *ms9*, *ms8*, *msp*, *ms3*, and *ms2* loci against themselves using BLASTX ($E < 10^{-6}$, Altschul et al., 1997). Custom perl scripts were used to parse the output into homologous groups of sequences ($E < 10^{-20}$, Graham et al., 2004). After removing splice variants, we were able to identify 144 groups with two or more genes per group. Once groups were identified, we used BLASTX ($E < 10^{-20}$, Altschul et al., 1997) against the Uniref100 protein database (Apweiler et al., 2004, version 06/2010) to annotate the putative function of each group.

Bioinformatic Identification of Male-Sterility Candidate Genes

Computational analysis using TBLASTX ($E < 10^{-6}$, Altschul et al., 1997) was used to compare known male-sterility genes from the model plant systems Arabidopsis (TAIR, www.arabidopsis.org) and rice (Rice Genome Annotation Project, rice.plantbiology.msu.edu). Sixteen sequences from rice (*Oryza sativa* Japonica) and 33

sequences from Arabidopsis were compared against all predicted genes in the soybean genome (Glyma 1, Schmutz et al., 2010). If the top soybean gene corresponded to a male-sterility locus, reciprocal TBLASTX analyses was performed against the entire genome of the original query sequence. Only if both genes identified each other as the best BLAST match in their respective genomes were the genes considered orthologs. BLASTX ($E<10^{-20}$, Altschul et al., 1997) also was used to identify male-sterility candidate genes by comparing Uniref100 protein database (Apweiler et al., 2004) against the genes within each male-sterility loci that had a significant overrepresentation of floral reads (*ms9*SF, *ms8*SF, *msp*SF, *ms3*SF, and *ms2*SF). Annotations were queried against available literature to identify candidate genes possibly involved in male sterility.

Identification of Homeologous Regions

To identify homeologous regions for each of the five male-sterility loci above, we first visualized each region using the SoyBase genome browser (www.soybase.org). We then downloaded the homeologous sequence and used BLASTN ($E < 10^{-4}$, Altschul et al., 1997) to confirm homology to the male-sterility regions. After obtaining the blast report, we used a locally installed version of the Artemis comparison tool (ACT, Carver et al., 2005) to visualize homeology between the male-sterility region and its homeolog. Homeolog sequences were trimmed, as needed, to match their respective male-sterility loci (Fig. 1-3).

Discerning Genes with Floral Expression

For analyses of all genes in the male-sterility regions (*ms9*A, *ms8*A, *ms3*A, *msp*A, *ms2*A, *ms9*HA, *ms8*HA, and *ms2*HA), every gene located between the flanking markers was included. Next, the gene IDs within each male-sterility locus or homeologous region were used to query RNA-Seq atlas (Severin et al., 2010, www.soybase.org/soyseq/). For every

gene, we recorded the number of raw floral and nonfloral tissue reads. We then divided the data into two separate floral classes: genes with at least one floral read (*ms9*F, *ms8*F, *ms3*F, *msp*F, *ms2*F, *ms9*HF, *ms8*HF, and *ms2*HF) and genes significantly overrepresented with floral reads (*ms9*SF, *ms8*SF, *ms3*SF, *msp*SF, *ms2*SF, *ms9*HSF, *ms8*HSF, and *ms2*HSF). Fisher's exact test was used to identify genes significantly overrepresented with floral reads by comparing read distribution in each gene relative to all genes in the genome (Fisher, 1949). To correct for oversampling, a Bonferoni correction was applied (P-value ≤ 0.05, Bonferroni, 1936). BLASTN ($E < 10^{-20}$, Altschul et al., 1997) analyses comparing male-sterility and male-sterility homeolog floral (F) and significantly floral (SF) datasets were used to determine if homeologous genes were maintained when floral expression data was used as a factor for parsing genes into datasets.

Cis-Element Over-Representation Analyses

Using custom perl scripts the sequence 1,000 base pairs (bp) upstream of the start codon (5' end of the gene) was extracted for each gene from Glyma1 genome assembly (Schmutz et al., 2010). To identify TFBS, we utilized two Clover analyses comparing each dataset to all promoters in the soybean genome sequence. First, we identified all TFBS presented within each dataset ($t = 1.0$, Frith et al., 2004) to establish a background distribution of TFBS. Second, we narrowed our focus to identify significantly overrepresented TFBS ($t \leq 0.05$, Frith et al., 2004). The Clover analyses were performed on a local bioinformatics machine, which identified the overrepresented TFBS from the curated TRANSFAC database (Release 7.0, Matys et al., 2006). To parse Clover output files, we used custom perl scripts to determine counts of overrepresented TFBS within each dataset, restricting ourselves to the 126 plant motifs (Tables 3 and 5).

RESULTS

Male-sterility Regions

In soybean, four male-sterility loci affected by environmental interactions have been identified: *ms9* (Wiebbecke et al., 2011), *ms8* (Perez-Sackett et al., 2011), *msp* (Stelley and Palmer, 1980), and *ms3* (Wiebbecke et al., unpublished data). These loci have been mapped to molecular linkage groups (MLG) N (chromosome Gm03, Cervantes-Martinez et al., 2007), M and D1b (chromosomes Gm07 and Gm02, Frasch et al., 2010), and D1b (chromosome Gm02, Cervantes-Martinez et al., 2009), respectively. The *ms2* locus, which is not sensitive to environmental cues, was mapped to MLG O (chromosome Gm10, Cervantes-Martinez et al., 2007). For each of these loci, publicly available simple sequence repeat (SSR) markers were used to map the male-sterility phenotype. By comparing the sequences of the nearest flanking SSR markers to the recently completed soybean genome sequence (Glyma 1, Schmutz et al., 2010) using BLASTN ($E < 10^{-20}$, Altschul et al., 1997), we were able to identify the genomic regions corresponding to each of these male-sterility loci (Table 1). Not surprisingly, each of the regions varied significantly in size, ranging from 63.9 kilobases (Kb) for *ms8* to 6.65 megabases (Mb) for *msp*. Similarly, while the *ms8* locus contained 14 predicted genes, *msp* contained 408 predicted genes.

Cluster Analysis of Male-Sterility Regions

Since little is known about the molecular basis of male sterility in soybean, our first objective was to compare genes across male-sterility loci to determine if common genes within these loci could play a role in male sterility. The 1,198 predicted genes (Glyma 1, Schmutz et al., 2010) within the *ms9*, *ms8*, *msp*, *ms3*, and *msp* loci were compared against themselves using TBLASTX ($E < 10^{-20}$, Altschul et al., 1997). Custom perl scripts were used

to cluster homologous sequences across loci. We identified 144 clusters containing two or more related genes (data not shown). BLASTX ($E<10^{-4}$, Altschul et al., 1997) analyses against the Uniref100 protein database (Apweiler et al., 2004) was used to annotate the putative function of each cluster. Only a single cluster was identified with any putative involvement in male sterility. This cluster contained 22 genes from the *ms9*, *msp*, *ms3*, and *ms2* regions with homology to a pentatricopeptide repeat (PPR) protein (NCBI reference sequence no. NM_122776.2). PPR genes are involved in diverse regulation of mitochondrial gene expression (Andrés et al., 2007) and PPR motifs have been identified in restorer genes in cytoplasmic male-sterile systems (Bentolila et al., 2002; Hanson and Bentolila, 2004). However, PPR genes are not associated with nuclear male sterility. Therefore these genes are unlikely to be involved in male sterility conferred by *ms9*, *ms8*, *msp*, *ms3*, and *ms2*.

Floral Expression within Male-Sterility Loci

Our previous analysis revealed little gene conservation across male-sterility loci. However, since male sterility affects pollen development, we could assume that candidates for male-sterility genes should be expressed in floral tissues. We used the RNA-Seq atlas (Severin et al., 2010, http://www.soybase.org/soyseq/), to obtain expression information for each of the genes within the male-sterility loci. Expression data in the RNA-Seq atlas was derived from plant material from the introgression line of *G. soja* (PI 468916) LG I protein QTL into *G. max* (A81-356022), which is not male sterile. Since the RNA-Seq atlas comes from a fertile line, we would expect all genes involved in floral development to be expressed normally, however it is important to note that genes with very specific floral expression patterns might not be represented. We extracted raw (unnormalized) expression data for each of the 1,198 genes in the *ms9*, *ms8*, *msp*, *ms3*, and *ms2* loci. We grouped expression data by

floral (open flowers) and nonfloral tissues [young leaf, one cm pod, pod shell 10 days after fertilization (DAF), pod shell 14 DAF, seed 10 DAF, seed 14 DAF, seed 21 DAF, seed 28 DAF, seed 35 DAF, seed 42 DAF, root, and nodule]. Since we assumed that candidate male-sterility genes would normally be expressed in floral tissue, we defined genes by their expression in floral tissues by creating two classes of candidate male-sterility genes (Table 1). The first class contained all genes with at least one floral read. In the largest locus, *ms2*, 91% of genes contained at least one floral read (designated floral, F). In contrast, only 55% of genes in the *msp* locus were expressed in flowers. The second class of floral genes contained genes that were significantly overrepresented (P-value ≤ 0.05) with floral reads relative to other sequenced tissues in the RNA-Seq atlas (Severin et al., 2010) and other genes in the genome. These genes, designated significantly floral (SF), were identified by performing a Fisher's exact test (Fisher, 1949) and correcting for over sampling (Bonferroni, 1936). The *ms8* locus had the highest percentage of significantly overrepresented floral genes with 43% (6 of 14) while *ms9* had the least with 23% (35 of 150). Using these two classes, we could potentially reduce the number of candidate genes from 1,198 to 908 with at least one floral read, which was further reduced to 324 genes with significant floral expression.

Bioinformatic Identification of Male-Sterility Candidate Genes

Our next step was to use a bioinformatics approach to determine if homologs of genes known to be involved in male sterility or pollen development in other systems were present in the soybean male-sterility loci. A complete survey of the available literature was performed to identify genes with known involvement in pollen development and/or male sterility. A recent review by Wilson and Zhang (2009) summarized all genes in Arabidopsis

and Rice involved in pollen development. In addition, publications by Liu et al. (2001), Yang et al. (2006), Woo et al. (2008), Yang et al. (2009), Quilichini et al. (2010), Xu et al. (2011), and Zhou et al. (2011) identified genes or candidate genes for thermosensitive male sterility. The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) and GenBank were queried for the 49 corresponding sequences. These sequences were compared to all predicted genes in the soybean genome using BLASTX (Altschul et al., 1997). In several cases, the best match was found within the known male-sterility loci, *ms3*, *msp*, and *ms2*. No homologs of known male-sterility genes were identified for either *ms9* or *ms8*. For the *ms3* locus, we identified Glyma02g11970 (1.0 E×10^{-114}), a homolog of rice *MS1* (*MALE STERILITY1*, Os09g27620), a plant homeodomain (PHD) transcription factor involved in regulation of late tapetal gene expression and pollen wall development (Ito et al., 2007; Wilson and Zhang, 2009). For the *msp* locus, we identified two possible candidates, Glyma02g40810 (E-value, 0.0) a homolog of rice UGPase (Os09g0553200), involved in conversion and transport of sugars for cell wall biosynthesis (Xiao et al., 2009) and Glyma02g41020 (2.0 E×10^{-67}) another homolog of rice *MS1*. Finally, in the *ms2* region, we identified a candidate gene, Glyma10g42830 (2.0 E×10^{-43}) a homolog of the rice *TDR* gene (*TAPETUM DEGENERATION RETARDATION*, Li et al., 2006) (NCBI Gene ID: XM_463907.1). Glyma10g42830 (1.0 E×10^{-120}) was identified again via BLASTX Arabidopsis query AT2G16910 (*AMS, ABORTED MICROSPORE*; Sorenson et al., 2003), which is orthologous to OsTDR. Mutants for *tdr*, in rice, and *ams*, in Arabidopsis, have shown delayed tapetal breakdown and programmed cell death. For the candidate male-sterility genes Glyma02g11970 (*ms3*), Glyma02g40810 and Glyma02g41020 (*msp*), and Glyma10g42830 (*ms2*), reciprocal TBLASTX (E < 10^{-20}, Altschul et al., 1997) against all

predicted genes in the Arabidopsis and rice genomes confirmed the sequences were most homologous to each other, further confirming their candidacy for male-sterility genes in soybean.

Since this approach failed to identify candidate genes for *ms8* or ms9, we broadened the approach to allow identification of additional candidate genes. Again, we utilized BLASTX ($E<10^{-20}$, Altschul et al., 1997) to compare the genes within each male-sterility locus that were significantly ($P < 0.05$) overrepresented with floral reads to the Uniref100 protein database (Apweiler et al., 2004, version 06/2010). Annotation information obtained by BLAST analyses was used to search the available literature for homologous genes known to be involved in pollen development and/or male sterility. Using this approach, we were able to identify additional candidate genes in all male-sterility loci except *ms8*. For the *ms9* locus, we identified three candidates: Glyma03g31520 (UniRef100_O24543, $3.0 \times E^{-70}$) and Glyma03g31530 (UniRef100_P13089, $1.0 \times E^{-107}$) both Auxin-Indole-3-acetic acid (Aux/IAA) responsive genes, which are both involved in thermosensitive male-sterility (Sakata et al., 2010), and Glyma03g31590 (UniRef100_B9S4T2, E-value = 0.0), a glucosyltransferase gene essential for pollen development (Xiao et al., 2009; Reyes et al., 2010). For *msp*, we identified six candidate genes: i) Glyma02g36700 (UniRef100_P26413, E-value = 0.0) a heat shock protein (HSP), similar to the HSP encoded by Arabidopsis *THERMOSENSITIVE MALE STERILE1* (*TMS1*, Yang et al., 2009), ii) Glyma02g37350 (UniRef100_B9RJR9, $1.0 \times E^{-108}$) encodes a flavonol synthase involved in flavonoid biosynthesis leading to floral pigmentation and attraction of pollinators (Schijlen et al., 2004), iii) Glyma02g38260 (UniRef100_Q8RVH8, $1.0 \times E^{-138}$) had homology to an Aux/IAA responsive gene, iv) Glyma02g39680 (UniRef100_B9T3Q2, $1.0 \times E^{-136}$) is a UDP-

glucose glucosyltransferase involved in thermosensitive male-sterility (Xiao et al., 2009; Reyes et al., 2010), v) Glyma02g40650 (UniRef100_B9SJM6, E-value = 0.0) is an Auxin response factor (ARF), which can be involved in various developmental processes including floral development (Kumar et al., 2011), and vi) Glyma02g40810 (UniRef100_Q8W557, E-value = 0.0), previously identified in the BLASTX comparison of male-sterility genes in rice. For the *ms3* locus, we identified ten genes: i) Glyma02g11060 (UniRef100_B9HIZ2, $1.0 \times E^{-117}$) and ii) Glyma02g13420 (UniRef100_Q533S8, $1.0 \times E^{-87}$) are *APETALA2* (*AP2*) and *AP1* domain containing transcription factors involved in floral organ identity (Melzer et al., 2009), iii) Glyma02g11530 (UniRef100_Q0N3X, $1.0 \times E^{-131}$) a Caffeoyl-CoA methyltransferase potentially involved in exine formation during pollen maturation (Xiao et al., 2009), iv) Glyma02g11610 (UniRef100_B9RYF1, $1.0 \times E^{-160}$), v) Glyma02g11670 (UniRef100_B9RYE0, $1.0 \times E^{-162}$), and vi) Glyma02g11680 (UniRef100_B9RYE0, $1.0 \times E^{-161}$), encoding UDP-glucose glycosyltransferases (Xiao et al., 2009), vii) Glyma02g12080 (UniRef100_C4B6D7, $2.0 \times E^{-82}$) had homology to Dof-like transcription factor involved in diverse biological processes and can be specifically expressed in flowers (Yanagisawa, 2004; Wang et al., 2007), viii) Glyma02g13390 (UniRef100_Q8GTF, $3.0 \times E^{-25}$) an AGL transcription factor involved in floral organ identity (Tilly et al., 1998), ix) Glyma02g13400 (UniRef100_A5YBS4, $8.0 \times E^{-22}$), *SEPALLATA1* (*SEP1*) also involved in floral organ identity (Melzer et al., 2009), and x) Glyma02g13790 (UniRef100_D0EWD, E-value = 0.0) a heat shock protein (HSP). Finally, for the *ms2* locus the technique helped identify two candidate genes: i) Glyma10g43850 (UniRef100_Q53B70, $1.0 \times E^{-117}$), a chalcone-flavonone isomerase also involved in flavonoid biosynthesis (Schijlen et al., 2004) and ii) Glyma10g44260 (UniRef100_A4GXC9, $1.0 \times E^{-156}$), a sugar transporter (Reyes et al., 2010;

Xiao et al., 2009). Using this combined bioinformatic approach, we identified 23 candidate genes that are putatively involved in male sterility in soybean. Further analyses of these genes can be prioritized based on best homology to known male-sterility genes in Arabidopsis and rice.

Cis-element Over-Representation Analyses of Male-Sterility Loci

We were particularly interested in the difference between environmentally-sensitive male sterility and environmentally-stable male sterility. Our hypothesis was that although both types of male sterility have floral phenotypes, the genes within each locus would have different functions and expression patterns. The timing of expression would be driven by transcription factors (TFs). In the case of environmentally-sensitive male-sterility loci, changes in environment could activate certain TFs, which would bind to transcription factor binding sites (TFBS) in the promoters of the male-sterility genes and alter their expression. Therefore, expression of different environmentally-sensitive loci could be regulated by related sets of TFs. While our data would not allow us to identify the transcription factors directly, we could analyze the promoters of genes within male-sterility loci to determine which TFBS were present. To do so, we used the same floral classes described previously: all genes with any floral expression in the locus (designated F) and all genes significantly ($P \leq 0.05$) overrepresented with floral reads in the locus (designated SF). In addition, we added a third class containing all genes in the locus (designated A). This class was included to determine the frequency of each TFBS within each region, regardless of expression. For each gene within each data set and each male-sterility locus, we extracted 1,000 base pairs (bp) of promoter sequence. We then used Clover (Frith et al., 2004) in conjunction with the TRANSFAC (version 7, Matys et al., 2006) transcription factor database to identify all

significantly ($t \leq 0.05$) overrepresented TFBS in each data set, relative to all promoters in the soybean genome (66,210 genes). The analysis was limited to the 126 plant TFs present in TRANSFAC. Once this analysis was completed, our primary objectives were to identify TFBS common across multiple environmentally sensitive male-sterility loci but absent in the environmentally stable locus, *ms2*. The results of this analysis could provide insight into the molecular basis of environmental sensitivity in male-sterility loci. Further identification of unique TFBS within each of the male-sterility loci could help explain the distinct pollen phenotypes for each male-sterility locus.

Prior to analyzing the male-sterility loci, we wanted to validate the approach of using TFBS to identify candidate genes. Therefore, we first used Clover analysis to identify TFBS significantly overrepresented ($t \leq 0.05$, Frith et al., 2004) in the previously identified soybean orthologs of Arabidopsis and rice male-sterility genes. All overrepresented TFBS identified in the promoter region of each of the four male-sterility candidate genes (Glyma02g11970 for *ms3*, Glyma02g40810 and Glyma02g41020 for *msp*, and Glyma10g42830 for *ms2*) are presented in Table 2. There were no TFBS shared across the soybean male-sterility orthologs, however, there were TFBS unique to the candidate genes involved in floral regulation. For Glyma02g11970, two overrepresented TFBS were identified in the *ms3*A dataset: AG_02 and AGL15_01, which are binding sites for the Agamous (AG) and AG-Like (AGL) transcription factors involved in floral development (Ito et al., 2004; Tripathi and Tuteja, 2007). For Glyma02g40810, five overrepresented TFBS were identified in *msp*F or *msp*SF datasets and not in the *msp*A dataset; of those, three TFBS bind to transcriptional regulators directly involved in floral development: i) auxin response factor binding site (ARF_Q2) where auxin response factors are involved in auxin-mediated responses during

floral development (Kumar et al., 2011), ii) GAMYB_Q2 (binds to a transcriptional activator, R2R3MYB protein, that interacts with Dof proteins in gibberellin (GA) signaling, Zou et al., 2008), and iii) indeterminate1 (ID1_01) a floral regulator protein regulating transition to flowering in maize (Wong and Colasanti, 2007). For Glyma10g42830, ten overrepresented TFBS were identified in the *ms2*A dataset. TFBS of particular interest were: GAMYB_01 (variant of GAMYB_Q2 identified in both *msp* orthologous male-sterility candidate genes), GT1_Q6 (binds to a pollen specific enhancer, Twell et al., 1991; Eyal et al., 1995), and WRKY_Q2 (binds WRKY TF involved in responding to biotic and abiotic stresses, Pandey and Somssich, 2009).

Upon validation of this approach, we analyzed each of the male-sterility loci and identified various TFBS significantly overrepresented ($t \leq 0.05$, Frith et al., 2004) in each dataset relative to all promoters in the genome. A comprehensive list of all TFBS for male-sterility loci is presented in Table 3. Clover analyses of the *ms9* locus identified 27 (*ms9*A), 25 (*ms9*F), and 8 (*ms9*SF) significant TFBS. For *ms8*, only a single overrepresented TFBS was identified for *ms8*A and *ms8*F, whereas two TFBS in *ms8*SF were significant. For the *msp* locus, Clover analyses found 41 (*msp*A), 48 (*msp*F), and 32 (*msp*SF) TFBS. The *ms3* locus contained 8 (*ms3*A), 10 (*ms3*F), and 6 (*ms3*SF) TFBS. Across the four environmentally sensitive male-sterility loci and three data classes, a total of 54 TFBS were overrepresented in at least one of the four loci. No specific TFBS was common across all environmentally sensitive male-sterility loci. However, all environmentally sensitive male-sterility loci did have at least one variation of the CArG binding site [AGAMOUS (AG), AG-Like (AGL), or AG-binding protein (AGP)], in which the transcription factors have been shown to be involved in regulation of floral development (Tilly et al., 1998). Interestingly,

the AGL3 TF, which is redundant to SEPALLATA and is involved in determining identity of floral organs (Melzer et al., 2009), and AGL15, has been shown to play a role in flower senescence (Tripathi and Tuteja, 2007), were not significantly overrepresented in the environmentally stable *ms2* locus. When we compared the TFBS found in the A, F, or SF data classes across male-sterility loci, 28 were unique to the F or SF classes. The plant-specific DNA-binding with one finger (Dof) TFBS was common across all male-sterility loci. Dof transcription factors are involved in a variety of biological processes (Yanagisawa, 2004), including light response (Yanagisawa and Sheen, 1998), plant defense (Zhang et al., 1995), glucosinolate biosynthesis (Skirycz et al., 2006), and auxin response (de Paolis et al., 1996). More specifically in soybean, seven flower and pod specific Dof-like proteins were identified (Wang et al., 2007). The connections to floral development indicate genes containing AG, AGL, AGP, and Dof TFBS are potential candidates for each male-sterility locus.

Within each environmentally sensitive male-sterility locus, we were able to identify unique overrepresented TFBS in the F or SF datasets, which may help explain their distinct male-sterile pollen phenotypes and sensitivity to environmental cues. The *ms9*F dataset had one uniquely overrepresented TFBS, MADSA_Q2, which binds a MADS-box TF known to play key roles in floral development (de Folter et al., 2005). The *ms8*F dataset had one uniquely overrepresented TFBS, DOF1_01. The *msp* locus had the most unique overrepresented TFBS, three in *msp*F: ARF_Q2, GAMYB_Q2, and ID1_01, as well as two in *msp*SF, abscisic acid-response factor (ABF1_03) that is cold-inducible and involved in regulating stress response (Kim, 2006), and again, GAMYB_Q2. Finally, the *ms3* locus had

no uniquely overrepresented TFBS. Additionally, *ms9*A, *ms8*A, or *msp*A contained no unique TFBS.

The non-environmentally sensitive *ms2* locus analyses contained more TFBS than the environmentally-sensitive male-sterility loci in each dataset: 43 (*ms2*A), 51 (*ms2*F), and 52 (*ms2*SF), which was unexpected since it was slightly smaller in gene number than the *msp* locus. Eighteen TFBS were overrepresented in *ms2*F or *ms2*SF but not identified in *ms2*A. While the *ms2* datasets did not have the AG binding site motifs mentioned earlier, there was one unique AGL TFBS overrepresented in *ms2*SF dataset, AGL2_01, with AGL being a regulator of floral development (Tilly et al., 1998). The *ms2* locus only had one type of Dof TFBS overrepresented in *ms2*F and *ms2*SF datasets: DOF3_01. This TFBS is found in a peptidase gene in rice (Yanagisawa, 2004) and also was found overrepresented in the *msp*F dataset. The *ms2* locus had four unique TFBS, the previously mentioned AGL2_01, ATHB5_01 [binds homeodomain leucine zipper, (HDZip) family of TFs, which are positive regulators of abscisic acid (ABA) response, Kim, 2006], WRKY_Q2, and ZAP1_01 (binds zinc-responsive transcriptional activator, a maize ortholog of *APETALA1*, Mena et al., 1995).

Identification of Male-Sterility Homeologous Regions

Soybean is recognized as an ancient palaeopolyploid with a duplicated genome (Schmutz et al., 2010). If male-sterility homeologous regions are present, genes within those regions may have evolved distinct functions, expression patterns, or become non-functional (Force et al., 2005, Yang et al., 2006). Therefore, we were interested in identifying the homeologs for each mapped male-sterility loci. Using the soybean genome browser (http://www.soybase.org), we identified homeologous regions (designated H) for two of the four environmentally sensitive male-sterility regions (*ms9*H and *ms8*H on chromosomes 19

and 8, respectively) and also the environmentally stable *ms2* locus (chromosome 20). The other two environmentally sensitive male-sterility loci had partial homology to multiple regions in the soybean genome: *ms3* was partially homologous to Gm01 and Gm07, whereas *msp* had three partial homologs on Gm10, Gm14, and Gm17 (Table 4). To date, these homeologous regions have no known sterility traits of any kind leading us to believe that the male-sterility genes in these regions have either been lost or specialization has occurred altering their expression.

For further analyses we chose to focus on those homeologous regions with complete homeology to the known male-sterility loci. To confirm homology, each male-sterility locus was compared to its homeolog using BLASTN ($E < 10^{-4}$, Altschul et al., 1997) and visualized using the Artemis comparison tool (ACT, Carver et al., 2005). The ends of putative homeologous regions were trimmed to match the corresponding male-sterility loci (*ms9H:* Gm19:41176292 to 42986270 and *ms8H*: Gm08:18310741 to 18484951, Fig. 1 and Fig. 2). In the case of *ms2,* Artemis visualization identified a 9.49 Mb nonhomeologous fragment separating two sections of the homeolog. The nonhomeologous section was removed and the two remaining fragments were treated as one complete homeolog (*ms2H,* Gm20:33040389 to 36501513; 45993589 to 46721826, Fig.3). Similar to the male-sterility loci, differences in size and number of genes within each homeologous region also were detected. The *ms9*HA and *ms8*HA datasets had 153 and 20 genes, respectively, similar to gene counts for their appropriate male-sterility regions. However, the *ms2*HA was 37% larger than the *ms2*A dataset (510 to 373 genes, respectively).

Floral Expression within Male-Sterility Homeologs

Similar to characterization of the male-sterility loci, we evaluated each male-sterility homeologous region for floral expression using the RNA-Seq atlas (Severin et al., 2010). The *ms2*HF, *ms9*HF, and *ms8*HF datasets had 403, 134, and 13 floral genes, respectively (Table 4). By limiting the analysis to significantly overrepresented floral genes, the *ms2*HSF, *ms9*HSF, and *ms8*HSF datasets had 110, 40, and seven floral genes. The *ms8*H decreased in both floral and significantly floral expression (65% and 35%, respectively) compared to *ms8* (78% and 42%, respectively). In contrast, the *ms9*H exhibited more floral expression than its respective male-sterility locus (*ms9*) for both floral and significantly floral classes (88% to 82% F; 26% to 23% SF, respectively). While the *ms2*HF decreased in floral expression compared to *ms2*F (79% to 91%), no difference was observed in significantly floral expression (*ms2*HSF, 26% of total genes) when compared to its respective male-sterility locus (*ms2*SF).

Similarity within Male-Sterility Homeologous Regions for Floral Expression

BLASTN ($E < 10^{-20}$, Altschul et al., 1997) was used to compare genes within male-sterility loci to genes within their respective homeologous region. For each comparison, we divided the genes by their expression class (F, SF). The greatest correspondence between genes and expression between homeologous regions was found using the F class. Comparisons of *ms2*F and *ms2*HF revealed that 78% of genes were shared between regions (317 of 406 *ms2*HF genes similar to *ms2*F). Similarly, 77% of genes were shared between *ms9*F and *ms9*HF data sets (106 of 137 *ms9*HF genes) and 69% of genes were shared between the *ms8*F to *ms8*HF data sets (9 of 13 *ms8*HF genes). When using the SF data sets, retention of genes and their expression decreased to less than 50%: *ms2*SF to *ms2*HSF 48% (53 of 110 *ms2*HSF), *ms9*SF to *ms9*HSF 48% (19 of 40 *ms9*HSF genes), and *ms8*SF to

*ms8*HSF at 43% (3 of 7 *ms8*HSF genes). These findings further support our hypothesis that subfunctionalization of homeologous regions may be regulated by changes in TFBS driving the differential expression patterns of homeologous genes.

To determine if the *ms9* and *ms2* candidate male-sterility genes we identified earlier were present in the homeologous regions, we used BLASTN ($E<10^{-4}$, Altschul et al., 1997) to compare Glyma03g31520, Glyma03g31530, and Glyma03g31590 (*ms9*) and Glyma10g42830, Glyma10g43850, and Glyma10g44260 (*ms2*) to the predicted genes in the *ms9*H and *ms2*H loci. The *ms8* and *ms8*H loci were not analyzed since no candidate genes were identified in an earlier analysis. For each gene, we were able to identify the corresponding homeolog (Glyma19g34370, Glyma19g34380, and Glyma19g34420 (*ms9*H); Glyma20g24170, Glyma20g38570, Glyma20g39040 (*ms2*H). The homeologs for Glyma03g31520, Glyma03g31530, Glyma03g31590 (*ms9*), and Glyma10g44260 (*ms2*) were present in the same (A, F, and/or SF) data classes. However, for the *ms2* and *ms2*H loci, Glyma10g42850 is significantly overrepresented with floral reads, while its homeolog (Glyma20g38570) was not. These data demonstrate the floral specificity of three candidate genes from the *ms9* locus and two of the candidate genes from the *ms2* locus, suggesting expression data can be used to prioritize genes for future analyses. Interestingly, Glyma10g42830 (*ms2*) and its homeolog Glyma20g24170 (*ms2*H), which showed greatest homology to the Arabidopsis *AMS* gene (*ABORTED MICROSPORE*, Sorenson et al., 2003), had very little evidence of expression in the RNA-Seq atlas. While Glyma10g42830 had six assigned reads (none floral), Glyma20g24170 had none. This suggests Glyma10g42830 has a very specific floral expression pattern not picked up in the RNA-Seq atlas or may not be a candidate gene for *ms2*.

Cis-element Over-Representation Analyses of Male-Sterility Homeologs

As before, we used Clover analyses to discover overrepresented TFBS ($t \leq 0.05$, Frith et al., 2004) within homeologs of male-sterility loci (*ms9*H, *ms8*H, and *ms2*H). A comprehensive list of all TFBS for male-sterility homeologs is presented in Table 5. Significantly overrepresented TFBS were identified for the homeologs of the environmentally sensitive male-sterility loci *ms9* and *ms8*. For *ms9*H, 21 TFBS were found in *ms9*HA, 22 in *ms9*HF, and 16 in *ms9*HSF. For *ms8*H, five TFBS were found in *ms8*HA, six in *ms8*HF, and eight in *ms8*HSF. A total of 26 different TFBS were found between *ms9*HA and *ms8*HA datasets. In total, in the *ms9*HF, *ms9*HSF, *ms8*HF, and *ms8*HSF datasets had 14 overrepresented TFBS that were not significant in their respective all genes data set. Two TFBS were uniquely overrepresented in the *ms9*H region (AG_02 and BCP1_Q2) and one (ARR10_01) in the *ms8*H region. Interestingly, BCP1 is an anther specific gene from *Brassica campestris* required for male fertility (Xu et al., 1995), whereas ARR10_01 binds an authentic response regulator that plays roles in cytokinin signaling (Ishida et al., 2008). Clover analyses of the homoelogous region for *ms2* found 31 TFBS overrepresented in *ms2*HA, 36 in *ms2*HF, and six in *ms2*HSF. TGA1A_Q2_01, which binds a transcriptional activator basic leucine zipper (bZIP) protein, was the only unique TFBS in the *ms2*H region, also identified in the *ms2*HA and *ms2*HF datasets. There was no overrepresented TFBS shared across all three homeologous regions.

Cis-element Over-Representation Analyses Between Male-Sterility Loci and Their Homeologs

Our final computational comparison was to evaluate each male-sterility region relative to its homeolog. When comparing *ms9* to *ms9*H, 23 TFBS were commonly

overrepresented across homeologous regions, seven of which were identified in the same classes for *ms9* and *ms9*H. 14 TFBS were significantly overrepresented in *ms9* but not significant in *ms9*H, some key TFBS were ABF_Q2, AG_01, AGL1_01, AGL1_02, AGL15_01, MADSA_Q2, PBF_01 a member of the Dof family (Yanagisawa and Schmidt, 1999), and RAV1_02, which binds to a cold-responsive transcription factor (Yamasaki et al., 2004), (Supplemental Table 1). Similarly, for the *ms8* locus compared to *ms8*H, one basic helix-loop-helix TFBS (BHLH66_01) was common to both regions, whereas 12 TFBS were differentially identified between *ms8* and *ms8*H. Two TFBS (AGP1_01 and DOF1_01) were unique to *ms8* compared with *ms8*H (Supplemental Table 2). Finally, in comparing *ms2* locus to *ms2*H, 38 TFBS were commonly overrepresented to both loci. However, only four were identified in the same data classes for *ms2* and *ms2*H (Supplemental Table 3). There were 28 TFBS differentially overrepresented between *ms2* and *ms2*H, 23 were uniquely overrepresented in the *ms2* locus compared with *ms2*H, some intriguing TFBS were ABF_Q2, AG_01 and_03, AGL1_02, AGL2_01, AGP1_01, BZIP910_02, BZIP911_01 and _02 (bZIP transcription factors are activators of histone gene expression during floral development, Martínez-García et al., 1998), DOF3_01, GAMYB_01, GT1_Q6, RAV1_02, and WRKY_Q2, which suggests the genes containing these TFBS may be involved in regulation of floral development (Supplemental Table 3).

DISCUSSION

Palmer et al. (2001) summarized soybean male-sterility loci by detailing the various stages of microgametogenesis affected by each male-sterility gene. The step in pollen production where breakdown occurs can result in diverse pollen phenotypes. One of our first

bioinformatic analyses performed was to determine if related genes were responsible for male sterility. We were unable to identify any candidate male-sterility genes shared across multiple loci. In retrospect, the subtle differences in pollen phenotype and abnormalities in microgametogenesis in each male-sterility locus (Palmer et al., 2001) suggested that shared genes across loci were unlikely. Nevertheless, using a bioinformatics approach we were able to identify 23 diverse candidate genes for *ms9*, *msp*, *ms3*, and *ms2* indicating that different genes are controlling male sterility at each locus. The *msp*, *ms3*, and *ms2* loci each contained a homolog of a male-sterility gene identified from the model organisms Arabidopsis or rice. Reciprocal best TBLASTX ($E < 10^{-20}$, Altschul et al., 1997) confirms that between genomes, these genes are most related to each other, further supporting their involvement in male sterility in soybean. In addition to searching for soybean homologs of known male-sterility genes, we also used BLAST (Altschul et al., 1997), in conjunction with the available literature, to identify 20 additional genes with links to floral development and male sterility. Comparison of these data with the phenotypic observations made by Palmer et al. (2001) may further help elucidate candidate male-sterility genes.

For the *ms9* locus, we were unable to identify homologs of known male-sterility genes. However, we identified three candidate genes based on BLAST (Altschul et al., 1997) and review of the literature. These genes encoded two AUX/IAA responsive genes and a glycosyltransferase. Each of these genes is a good candidate for the *ms9* thermosensitive genic male-sterile (TGMS) phenotype (Wiebbecke et al., 2011), because glycosyltransferase and auxin response genes are involved in TGMS (Xiao et al., 2009; Sakata et al., 2010). The TGMS conferred by the glycosyltransferase and auxin biosynthesis genes resulted in complete male sterility in rice, Arabidopsis, and barley. However, the *ms9* male-sterile

phenotype occurs late in microgametogenesis, allowing production of some functional pollen, as well as abnormal non-functional pollen (Wiebbecke et al., 2011). It is possible these candidate genes are involved in male-sterility for *ms9* locus because they are thermosensitive, but they may have different TFBS altering detection of environmental cues.

In the case of *msp*, BLAST analyses using known male-sterility genes from model plants identified a gene homologous to the *ms-h* gene encoding a UGPase (Woo et al., 2008). In rice, UGPase silenced or cosupressed plants had pale, shrunken anthers, resulting in complete sterility, and were thermosensitive (Chen et al., 2007, 2009). Six additional candidate genes were identified by BLAST (Altschul et al., 1997) and review of the male-sterility literature: two Aux/IAA responsive genes, a UDP-glucose-glycosyltransferase (Xiao et al., 2009; Reyes et al., 2010), a heat shock protein (HSP, Yang et al., 2009), and flavonol biosynthesis gene (Schijlen et al., 2004) genes were identified, as well as the previously identified UGPase. For the *msp* locus, pollen degeneration can occur at any stage in development allowing some functional pollen to be produced resulting in a partial male-sterile phenotype (Palmer et al., 2001). The Aux/IAA, UDP-glucose-glycosyltransferase, and UGPase genes are influenced by temperature (Chen et al., 2007, 2009; Xiao et al., 2009; Sakata et al., 2010). Like the *ms9* locus, different regulatory elements in the promoter regions of these genes may alter their expression in the *msp* background resulting in partial-male sterility. The HSP gene in Arabidopsis (*TMS1*) and flavonol biosynthesis genes effect pollen-tube germination resulting in decreased male fertility (Schijlen et al., 2004; Yang et al., 2009). Further microscopic investigation is needed to determine if pollen-tube germination is effected in the *msp* male-sterile background.

One of the candidate genes identified for *ms3* was an ortholog of the rice *MALE STERILITY1* (*MS1*) gene encoding a plant homeodomain transcription factor (Ito et al., 2007). *MS1* mutants in Arabidopsis have shown variable pollen phenotype from semisterile (Ito et al., 2007) to no viable pollen (Wilson et al., 2001). Both pollen phenotypes are caused by altered tapetum degeneration and abnormal programmed cell death occurring after microspore release from the tetrad (Ito et al., 2007; Wilson and Zhang, 2009). In the case of *ms3*, abnormal tapetum degeneration appears to inhibit microspore release resulting in both viable and nonfunctional pollen (Palmer et al., 2001). In addition, *ms3* may have an immature pollen wall causing degeneration of microspores (Palmer et al., 2001). By using BLAST (Altschul et al., 1997) and reviewing the literature, ten additional candidate genes were identified within six functional groups: AP2, Caffeyol-CoA methyltransferase, UDP-glucose related, MADS-box TF, and HSP. Of these candidate genes, the UDP-glucose related, Caffeyol-CoA methyltransferase, and HSP are the most likely candidate genes because they previously have been shown to be involved in TGMS (Chen et al., 2007, 2009; Xiao et al., 2009; Yang et al., 2009).

For the *ms2* loci, we were able to identify two candidate genes from BLASTX of Arabidopsis and rice male-sterility genes. Again, an ortholog of the rice *MS1* gene was identified. The second candidate gene identified for *ms2* was an ortholog of *TAPETUM DEGENERATION RETARDATION* (*TDR*) in rice, and the *ABORTED MICROSPORE (AMS*) Arabidopsis gene. For both of these genes pollen development fails because of premature tapetum and microspore degeneration, typically just after microspore release from the tetrad (Sorenson et al., 2003; Li et al., 2006). Identification of these candidate genes is surprising because *ms2* male sterility is thought to be caused by an earlier failure during

microgametogenesis, causing degradation of microspores in the tetrad prior to microspore release, resulting in complete male sterility (Palmer et al., 2001). Based on gene annotation, we were able to identify two additional candidate genes, a chalcone-flavonol biosynthesis gene and sugar transporter both involved in floral and pollen development (Schijlen et al., 2004; Xiao et al., 2009; Reyes et al., 2010).

Each locus has a unique male-sterile pollen phenotype that may be the result of different genes. However, environmentally sensitive male-sterility loci may share common transcription factors regulating response to biotic cues for male-sterility genes. Our results suggest that TFs binding the AG, AGL, AGP, and Dof TFBS, identified in the environmentally sensitive male-sterility loci, may be involved in regulating those signaling cascades. In contrast, *ms2* which has no demonstrated environmental sensitivity, may be regulated by different TF classes. Identification of different transcriptional regulators in soybean male-sterility loci suggests different mechanisms influence male-sterility gene perception of environmental cues.

Our initial analysis began with the sequences of markers used to map male-sterility genes. Our bioinformatic analyses demonstrates that researchers can use computational methods to identify candidate genes from mapping data using the genomic and expression data readily available for soybean. The recently released soybean genome sequence (Glyma1, Schmutz et al., 2010), the soybean genome browser (www.soybase.org), and RNA-Seq atlas (Severin et al., 2010) are all available to the soybean community and comparative analyses can be made between soybean and other model species to identify candidate genes from other species.

The candidate genes we have identified in the bioinformatic analysis can be further investigated to see if they are responsible for male sterility at each locus. A more traditional molecular approach would be to design new single nucleotide polymorphism (SNP) or SSR markers in or near the candidate gene(s), which could then be used to screen a segregating population to determine if the marker segregates with the male-sterile phenotype. A more sophisticated method would be to use virus induced gene silencing (VIGS, Zhang et al., 2009) or RNAi (Libault et al., 2009) to knock down the candidate gene expression and then screen for the male-sterile phenotype.

Once a male-sterility gene is confirmed in soybean, it would enable breeders to develop and use SNP or SSR markers for marker assisted selection (MAS) to identify male-sterile, female-fertile plants. Developing a MAS system would eliminate one limitation mentioned by Palmer et al. (2001), described a requirement for successful implementation of hybrid seed production, "a selection system to identify 100% female plants". A more comprehensive view of environmental male sterility and the signaling cascade regulating response to biotic cues could also lead to the elimination of another constraint to hybrid soybean production, which is "a stable-uniform male-sterile, female-fertile plant to serve as the female parent in hybrid seed production" (Palmer et al., 2001). Developing a better understanding of the transcription factors regulating male-sterility and the TFBS within the promoters of male-sterility genes will enable the improvement of male-sterile systems in soybean and could ultimately lead to a broad use of hybrids in soybean cultivar development.

REFERENCES

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Andrés, C., C. Lurin, and I.D. Small. 2007. The multifarious roles of PPR proteins in plant mitochondrial gene expression. Physiol. Plant. 129:14-22.

Apweiler, R., A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, and L.S.L. Yeh. 2004. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 32:D115-D119.

Bentolila, S., A. A. Alfonso, and M.R. Hanson. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. PNAS 99:10887-10892.

Bonferroni, C.E. 1936. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 8:3-62.

Carver, T.J., K.M. Rutherford, M. Berriman, M.A. Rajandream., B.G. Barrell, and J. Parkhill. 2005. ACT: the Artemis comparison tool. Bioinformatics 21:3422-3423.

Caviness, C.E., H.J. Walters, and D.L. Johnson. 1970. A partially male-sterile strain of soybean. Crop Sci. 10:107-108

Cervantes-Martinez, I., D. Sandhu, M. Xu, E. Ortiz-Perez, K. K. Kato, H. T. Horner, and R.G. Palmer. 2009. The male sterility locus *ms3* is present in a fertility controlling gene cluster in soybean. J. Hered. 100:565-570.

Cervantes-Martinez, I., M. Xu, L. Zhang, Z. Huang, K.K. Kato, H.T. Horner, and R.G. Palmer. 2007. Molecular mapping of male sterility loci *ms2* and *ms9* in soybean. Crop Sci. 47:374-379.

Chen, R., Y. Pan, Y. Wang, L. Zhu, and G. He. 2009. Temperature-sensitive splicing is an important molecular regulation mechanism of thermosensitive genic male sterility in rice. Chinese Sci Bull. 54:2354-2362.

Chen, R., X. Zhao, Z. Shao, Z. Wei, Y. Wang, L. Zhu, J. Zhao, M. Sun, R. He, and G. He. 2007. Rice *UDP-Glucose Pyrophosphorylase1* is essential for pollen callose deposition and its cosupression results in a new type of thermosensitive gene male sterility. Plant Cell 19:847-861.

Eyal, Y., C. Curie, and S.McCormick. 1995. Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. Plant Cell 7:373-384.

Fisher, R.A. 1949. A preliminary linkage test with agouti and undulated mice; the fifth linkage-group. Heredity 3:229-241.

de Folter, S., R.G.H. Immink, M. Kieffer, L. Pařenicová, S.R. Henz, D. Weigel, M. Busscher, M. Kooiker, L. Colombo, M.M. Kater, B. Davies, and G.C. Angenent. 2005. Comprehensive interaction map of the Arabidopsis MADS box transcription factors. Plant Cell 17:1424-1433.

Force, A., W.A. Cresko, F.B. Pickett, S.R. Proulx, C. Amemiya, and M. Lynch. 2005. The origin of subfunctions and modular gene regulation. Genetics 170:433-446.

Frasch, R.M., C. Weigand, P.T. Perez, R.G. Palmer, and D. Sandhu. 2011. Molecular mapping of 2 environmentally sensitive male-sterile mutants in soybean. J. Hered. 102:11-16.

Frith, M.C., Y. Fu, L. Yu, J. Chen, U. Hansen, and Z. Weng. 2004. Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res. 32:1372-1381.

Goldberg, R.B., T.P. Beals, and P.M. Sanders. 1993. Anther development: basic principles and practical applications. Plant Cell 5:1217-1229.

Graham, M.A., K.A.T. Silverstein, S.B. Cannon, and K.A. VandenBosch. 2004. Computational identification and characterization of novel genes from legumes. Plant Physiol. 135:1179-1197.

Hadley, H.H., and W.J. Starnes. 1964. Sterility in soybeans caused by asynapsis. Crop Sci. 4: 421-424.

Hanson, M.R., and S. Bentolila. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:S154-S169.

Horner, H.T., and R.G. Palmer. 1995. Mechanisms of genic male sterility. Crop Sci. 35:1527-1535.

Ishida, K., T. Yamashino, A. Yokoyama, and T. Mizuno. 2008. Three type-B response regulators, *ARR1*, *ARR10* and *ARR12*, play essential but redundant roles in cytokinin signal transduction throughout the life cycle of *Arabidopsis thaliana*. Plant Cell Physiol. 49:47-57.

Ito, T., N. Nagata, Y. Yoshiba, M. Ohme-Takagi, H. Ma, and K. Shinozaki. 2007. Arabidopsis *MALE STERILITY1* encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell 19:3549-3562.

Ito, T., F. Wellmer, H. Yu, P. Das, N. Ito, M. Alves-Ferreira, J.L. Riechmann, and E.M. Meyerowitz. 2004. The homeotic protein AGAMOUS controls microsporogenesis by regulation of *SPOROCYTELESS*. Nature 430:356-360.

Jan. C.C. 1974. Genetic male sterility in wheat (*Triticum aestivum* L.): expression, stability, inheritance, and practical use. Ph.D. thesis. Univ. of California, Davis, Calif., USA.

Jenkins, J.N., J.C. McCarty, J. Wu, and O.A. Gutierrez. 2009. Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. Euphytica 167:397-408.

Kim, S.Y. 2006. The role of ABF family bZIP class transcription factors in stress response. Physiol. Plant. 126:519-527.

Kumar, R., A.K. Tyagi, and A.K. Sharma. 2011. Genome-wide analysis of *auxin response factor* (*ARF*) gene family from tomato and analysis of their role in flower and fruit development. Mol. Genet. Genomics 285:245-260.

Li, N., D.S. Zhang, H.S. Liu, C.S. Yin, X.X. Li, W.Q. Liang, Z.Yuan, B. Xu, H.W. Chu, J. Wang, T.Q. Wen, H. Huang, D. Luo, H. Ma, and D.B. Zhang. 2006. The rice *Tapetum Degeneration Retardation* gene is required for tapetum degradation and anther development. Plant Cell 18:2999-3014.

Li, J., and L. Yuan. 2000. Hybrid rice: Genetics, breeding, and seed production. Plant Breed. Rev. 17:15-158.

Libault, M., T. Joshi, K. Takahashi, A. Hurley-Sommer, K. Puricelli, S. Blake, R.E. Finger, C.G. Taylor, D. Xu, H.T. Nguyen, and G. Stacey. 2009. Large-scale analysis of putative soybean regulatory gene expression identifies a *Myb* gene involved in soybean nodule development. Plant Physiol. 151:1207-1220.

Liu, N., Y. Shan, F.P. Wang, C.G. Xu, K.M. Peng, X.H. Li, and Q. Zhang. 2001. Identification of an 85-kb DNA fragment containing *pms1*, a locus for photoperiod-sensitive genic male sterility in rice. Mol. Genet. Genomics 266:271-275.

Martínez-García, J.F., E. Moyano, M.J.C. Alcocer, and C. Martin. 1998. Two bZIP proteins from *Antirrhinum* flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. Plant J. 13:489-505.

Matys, V., O.V. Kel-Margoulis, E. Fricke, I. Liebich, S.Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A.E. Kel, and E. Wingender. 2006. TRANSFAC and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34:D108-D110.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004a. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. Crop Sci. 44:1226-1230.

McCarty, J.C., J.N. Jenkins, and J. Wu. 2004b. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: II. Genetic effects and genotypic values. Crop Sci. 44:1231-1235.

Melzer, R., W. Verelst, and G. Theiβen. 2009. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes *in vitro*. Nucleic Acids Res. 37:144-157.

Mena, M., M.A. Mandel, D.R. Lerner, M.F. Yanofsky, and R.J. Schmidt. 1995. A characterization of the *MADS-box* gene family in maize. Plant J. 8:845-854.

Ortiz-Perez, E., H.T. Horner, S.J. Hanlin, and R.G. Palmer. 2006. Insect-mediated seed-set evaluation of 21 soybean lines segregating for male sterility at 10 different loci. Euphytica 152:351-360.

Owen, F.V. 1928. A sterile character in soybeans. Plant Physiol. 3:223-226.

Palmer, R.G., P.T. Perez, E. Ortiz-Perez, F. Maalouf, and M.J. Suso. 2009. The role of crop-pollinator relationships in breeding for pollinator-friendly legumes: from a breeding perspective. Euphytica 170:35-52.

Palmer, R.G., J. Gai, H. Sun, and J.W. Burton. 2001. Production and evaluation of hybrid soybean. Plant Breed. Rev. 21:263-307.

Pandey, S.P., and I.E. Somssich. 2009. The role of WRKY transcription factors in plant immunity. Plant Physiol. 150:1648-1655.

de Paolis, A., S. Sabatini, L. De Pascalis, P. Costantino, and I. Capone. 1996. A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. Plant J. 10:215-223.

Pereira, T.N.S., N.R. Lersten, and R.G. Palmer. 1997. Genetic and cytological analyses of a partial-female-sterile mutant (*PS-1*) in soybean (*Glycine max*: Leguminosae). Am. J. Bot. 84:781-791.

Perez-Sackett, P.T., and R.G. Palmer. 2011. Effect of day and night temperature on the expression of male sterility of nuclear male-sterilt (*ms8ms8*) soybean line. Submitted to Euphytica.

Quilichini, T.D., M.C. Friedmann, A.L.Samuels, and C.J. Douglas. 2010. ATP-binding cassette transporter *G26* is required for male fertility and pollen exine formation in Arabidopsis. Plant Physiol. 154:678-690.

Reyes, F., G. León, M. Donoso, F. Brandizzi, A.P.M. Weber, and A. Orellana. 2010. The nucleotide sugar transporters *AtUTr1* and *AtUTr3* are required for the incorporation of UDP-glucose into the endoplasmic reticulum, are essential for pollen development and are needed for embryo sac progress in *Arabidopsis thaliana.* Plant J. 61:423-435.

Rice Genome Annotation Project. [http:// rice.plantbiology.msu.edu/].

Sakata, T., T. Oshino, S. Miura, M. Tomabechi, Y. Tsunaga, N. Higashitani, Y. Miyazawa, H. Takahashi, M. Watanabe, and A. Higashitani. 2010. Auxins reverse plant male sterility caused by high temperatures. PNAS 107:8569-8574.

Saxena, K.B., R. Sultana, N. Mallikarjuna, R.K. Saxena, R.V. Kumar, S.L. Sawargaonkar, and R.K. Varshney. 2010. Male sterility systems in pigeonpea and their role in enhancing yield. Plant Breed. 129:125-134.

Schijlen, E.G.W.M., C.H. Ric de Vos, A.J. van Tunen, and A.G. Bovy. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Scott, R.J., M. Spielman, and H.G. Dickinson. 2004. Stamen structure and function. Plant Cell 16:S46-S60.

Severin, A.J., J.L. Woody, Y. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Spect, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, and R.C. Shoemaker. 2010. RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol. 10:160-176.

Sharma R.K., and E. Reinbergs. 1976. Male sterility genes in barley and their sensitivity to light and temperature intensity. Ind. J. Genet. Plant Breed. 2:408-410.

Skirycz, A., M. Reichelt, M. Burow, C. Birkemeyer, J. Rolcik, J. Kopka, M.I. Zanor, J. Gershenzon, M. Strnad, J. Szopa, B. Mueller-Roeber, and I. Witt. 2006. DOF transcription factor *AtDof1.1* (*OBP2*) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. Plant J. 47:10-24.

Sorensen, A.M., S. Kröber, U.S. Unte, P. Huijser, K. Dekker, and H. Saedler. 2003. The Arabidopsis *ABORTED MICROSPORE* (*AMS*) gene encodes a MYC class transcription factor. Plant J. 33:413-423.

SoyBase. [http://www.soybase.org].

Stelly, D.M., and R.G. Palmer. 1980. A partially male-sterile mutant line of soybeans, *Glycine max* (L.) Merr.: Characterization of the *msp* phenotype variation. Euphytica 29:539-546.

Sun, Z.X., S.K. Min, and Z.M. Xiong. 1989. A temperature sensitive male sterile line found in rice. Rice Genet. Newsl. 6:116-117.

TAIR. [http://www.arabidopsis.org].

Tang, Z., L. Zhang, D. Yang, C. Zhao, and Y. Zheng. 2011. Cold stress contributes to aberrant cytokinesis during male meiosis I in wheat thermosensitive genic male sterile line. Plant Cell Environ. 34:389-405.

Tilly, J.J., D.W. Allen, and T. Jack. 1998. The CArG boxes in the promoter of the Arabidopsis floral organ identity gene *APETALA3* mediate diverse regulatory effects. Development 125:1647-1657.

Tripathi, S.K., and N. Tuteja. 2007. Integrated signaling in flower senescence: An overview. Plant Signal. Behav. 2:437-445.

Twell, D., J. Yamaguchi, R.A. Wing, J. Ushiba, and S. McCormick. 1991. Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. Genes Dev. 5:496-507.

Virmani, S.S., Z.M. Sun, T.M. Mou, A.J. Ali, and C.X. Mao. 2003. Two-line hybrid rice seed production, p. 53-64. *In* B. Hardy (ed.) Two-line hybrid rice breeding manual. International Rice Research Institute. Los Baños, Philippines.

Wang, H.W., B, Zhang, Y.J. Hao., J. Huang, A.G. Tain, Y. Liao, J.S. Zhang, and S.Y. Chen. 2007. The soybean *Dof-type* transcription factor genes, *GmDof4* and *GmDof11*, enhance lipid content in the seeds of transgenic Arabidopsis plants. Plant J. 52:716-729.

Wiebbecke, C.E., M.A. Graham, S.C. Cianzio, and R.G. Palmer. 2011. Day temperature influences the male-sterility locus *ms9* in soybean. Submitted to Crop Sci.

Wilcock, C., and R. Neiland. 2002. Pollination failure in plants: why it happens and when it matters. Trends Plant Sci. 7:270-277.

Wilson, Z.A., and D.B. Zhang. 2009. From Arabidopsis to rice: pathways in pollen development. J. Exp. Bot. 60:1479-1492.

Wilson, Z.A., S.M. Morroll, J. Dawson, R. Swarup, and P.J. Tighe. 2001. The Arabidopsis *MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J. 28:27-39.

Wong, A.Y.M., and J. Colasanti. 2007. Maize floral regulator protein INDETERMINATE1 is localized to developing leaves and is not altered by light or the sink/source transition. J. Exp. Bot. 58:403-414.

Woo, M.O., T.H. Ham, H.S. Ji, M.S. Choi, W. Jiang, S.H. Chu, R. Piao, J.H. Chin, J.A. Kim, B.S. Park, H.S. Seo, N.S. Jwa, S. McCouch, and H.J. Koh. 2008. Inactivation of the *UGPase1* gene causes genic male sterility and endosperm chalkiness in rice (*Oryza sativa* L.). Plant J. 54:190-204.

Xi, D.W., W.J. Cheng, D.L. Yi, D.R. Yu, Z.L. Ning, X.X. Deng, and M. Li. 1997. Genetic analysis on the TGMS line in rape of B. napus Xiangyou 91S. In: R.L. Lu, X.B. Cao, F.M. Liao, and Y.Y. Xin, editors. Proceedings of the International Symposium on Two-Line Systems of Heterosis Breeding in Crops, 6-8 September 1997, Changsha, China. P 215-220.

Xiao, X., Y. Yang, Y. Yang, J. Lin, D. Tang, and X. Liu. 2009. Comparative analysis of young panicle proteome in thermo-sensitive genic male-sterile rice Zhu-1S under sterile and fertile conditions. Biotech Lett. 31:157–161.

Xu, H., R.B. Knox, P.E. Taylor, and M.B. Singh. 1995. *Bcp1*, a gene required for male fertility in Arabidopsis. PNAS 92:2106-2110.

Xu, J., C. Yang, Z. Yuan, D. Zhang, M.Y. Gondwe, Z. Ding, W. Liang, D. Zhang, and Z.A. Wilson. 2010. The *ABORTED MICROSPORES* regulatory network is required for postmeiotic male reproductive development in *Arabidopsis thaliana*. Plant Cell 22:91-107.

Xu. X., J. Bian, S. Liu, H. Song, N. Shi, Y. Tao, and H. Wong. 2011. Flower-specific expression of Arabidopsis *PCS1* driven by *AGAMOUS* second intron in tobacco decreases the fertility of transgenic plants. Mol. Breed. 27:337-346.

Yamasaki, K., T. Kigawa, M. Inoue, M. Tateno, T. Yamasaki, T. Yabuki, M. Aoki, E. Seki, T. Matsuda, Y. Tomo, N. Hayami, T. Terada, M. Shirouzu, T. Osanai, A. Tanaka, M. Seki, K. Shinozaki, and S. Yokoyama. 2004. Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor *RAV1*. Plant Cell 16:3448-3459.

Yanagisawa, S. 2004. Dof domain proteins: plant specific transcription factors associated with diverse phenomenon unique to plants. Plant Cell Physiol. 45:384-391.

Yanagisawa, S., and R.J. Schmidt. 1999. Diversity and similarity among recognition sequences of *Dof* transcription factors. Plant J. 17:209-214.

Yanagisawa, S., and J. Sheen. 1998. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell 10:75-89.

Yang, J.H., S.J. Han, E.K. Yoon, and W.S. Lee. 2006. Evidence of an auxin signal pathway, microRNA167-ARF8-GH3, and its response to exogenous auxin in cultured rice cells. . Nucleic Acids Res. 34:1892-1899.

Yang, K.Z., C. Xia, X.L. Liu, X.Y. Dou, W. Wang, L.Q. Chen, X.Q. Zhang, L.F. Xie, L. He, X. Ma, and D. Ye. 2009. A mutation in *THERMOSENSITIVE MALE STERILITY1*, encoding a heat shock protein with DnaJ and PDI domains, leads to thermosensitive gametophytic male sterility in Arabidopsis. Plant J. 57:870-882.

Yang, X., G.A. Tuskan, and M.Z-M. Cheng. 2006. Divergence of the *Dof* gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. Plant Physiol. 142:820-830.

Zhang, B., W. Chen, R.C. Foley, M. Buttner, and K.B. Singh. 1995. Interactions between distinct types of DNA binding proteins enhance binding to *ocs* element promoter sequences. Plant Cell 7:2241-2252.

Zhang, C., C. Yang, S.A. Whitham, and J.H. Hill. 2009. Development and use of efficient DNA-based viral gene silencing vector for soybean. MPMI 22:123-131.

Zhou, Y.F., X.Y. Zhang, and Q.Z. Xue. 2011. Fine mapping and candidate gene prediction of the photoperiod and thermo-sensitive genic male sterile gene *pms1(t)* in rice. J. Zhejiang Univ.-Sci.B (Biomed. and Biotech.) 12:436-447.

Zou, X., D. Neuman, and Q.J. Shen. 2008. Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellins signaling in aleurone cells. Plant Physiol. 148:176-186.

Table 1. Molecularly mapped male-sterility loci in soybean.

Male-sterility region/ Strain†	All tissues‡	Floral‡	Predicted genes	Predicted genes with > 1 floral reads	Predicted genes with significantly floral expression§	Chromosome number and location (bp)	Region size (Kb)	MLG	Flanking Markers	Gene range from Glyma1 genome assembly¶	Reference
ms3 T284H	196,297	16,219	253	206	86	Gm02:9239800 ..12976559	3736.8	D1b	Satt157 to Satt542	Glyma02g11060 to 02g14380	Cervantes-Martinez et al., 2009
ms8 T358H	15,129	1,410	14	11	6	Gm07:01 ..163911	63.9	M	Sat_389 to telomere	Glyma07g00200 to 07g00370	Frasch et al., 2010
ms9 T359H	1,323,458	7,094	150	123	35	Gm03:38596702 ..40192601	1595.9	N	Satt521 to Satt237	Glyma03g30740 to 03g32450	Cervantes-Martinez et al., 2007
msp T271H	291,761	23,643	408	227	99	Gm02:39888952 ..46538883	6649.9	D1b	Satt350 to Sat_069/ Satt172	Glyma02g35180 to 02g41380	Frasch et al., 2010
ms2 T375H T360H	328,265	25,283	373	341	98	Gm10:47601155 ..50969635	3368.5	O	Sat_109 to telomere	Glyma10g40090 to 10g44630	Cervantes-Martinez et al., 2007
Genome All	8,143,671	495,595	66,210	37,671		-	1,115,000	-	-	-	Schmutz et al., 2010

† T-strains with H suffix (e.g. T284H) the allele is carried as the heterozygote because the homozygote is sterile.

‡ Raw expression counts from 14 tissue types including root, nodule, young leaf, flower, pod, and seed with different time points during pod and seed development, RNA-Seq atlas (Severin et al., 2010).

§ Significantly floral genes identified by Fisher's exact test (Fisher, 1949) comparing read distribution for each gene relative to the soybean genome. A Bonferroni correction was applied for over-sampling (P-value ≤ 0.05, Bonferroni, 1936).

¶ Glyma 1, Schmutz et al. (2010) and SoyBase (www.soybase.org).

Table 2. Overrepresented ($t \leq 0.05$) plant transcription factor binding sites (TFBS) found in the promoter region of Arabidopsis and rice male-sterility gene orthologs were identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006). The dataset identifier is in parenthesis below each Glyma 1 assembly ID (Schmutz et al., 2010) in the heading.

Plant TFBS	**Glyma02g11970 (*ms3*A)**	**Glyma02g40810 (*msp*A)**	**Glyma02g40810 (*msp*F)**	**Glyma02g40810 (*msp*SF)**	**Glyma02g41020 (*msp*A)**	**Glyma02g41020 (*msp*F)**	**Glyma10g42830 (*ms2*A)**
ABF_Q2	0	0	0	0	1*	1*	0
ABZ1_01	0	0	0	0	3 *	3 *	0
AG_02	1 *	0	0	0	2 *	2 *	0
AGL15_01	1 *	0	0	0	2 *	2 *	2
ALFIN1_Q2	2	3 *	3 *	3 *	2 *	2 *	1
ARF_Q2	2	2	2 *	2	1	1 *	1
ATMYB15_01	1	0	0	0	0	0	1 *
CG1_Q6	1	3 *	3 *	3 *	2 *	2 *	0
CPRF2_01	0	0	0	0	1 *	1 *	0
CPRF2_Q2	0	0	0	0	1 *	1 *	0
CPRF3_01	0	0	0	0	1 *	1 *	0
CPRF3_Q2	0	0	0	0	1 *	1 *	0
DOF_Q2	1	2 *	2 *	2	0	0	1
DOF2_01	0	0	0	0	2	2 *	2
DOF3_01	0	0	0	0	4	4 *	2
E2F_Q2	0	0	0	0	2 *	2 *	0
EMBP1_02	0	0	0	0	2 *	2 *	1
GAMYB_01	0	1	1	1	3	3	2 *
GAMYB_Q2	0	1	1 *	1 *	2	2 *	1
GBF_Q2	0	2 *	2 *	2 *	0	0	0
GBP_Q6	0	1 *	1 *	1 *	0	0	0
GT1_Q6	0	0	0	0	0	0	1 *
HBPA1_Q6_01	0	1 *	1 *	1	1 *	1 *	0

Table 2. (continued)

Plant TFBS	Glyma02g11970 (*ms3*A)	Glyma02g40810 (*msp*A)	Glyma02g40810 (*msp*F)	Glyma02g40810 (*msp*SF)	Glyma02g41020 (*msp*A)	Glyma02g41020 (*msp*F)	Glyma10g42830 (*ms2*A)
ID1_01	0	1	1 *	1	2	2 *	0
KNOX3_01	0	2 *	2 *	2 *	1 *	1 *	1 *
LIM1_01	1	2 *	2 *	2 *	3 *	3 *	2 *
MYBAS1_01	0	1	1 *	1 *	4	4 *	0
NAC691_01	0	2 *	2 *	2 *	0	0	0
O2_02	1	0	0	0	2 *	2 *	0
O2_Q2	1	0	0	0	3 *	3 *	0
OCSBF1_01	1	1 *	0	1 *	0	0	0
OSBZ8_Q6	0	2 *	0	2 *	2 *	2 *	0
PBF_01	1	0	0	0	2 *	2 *	1
PCF2_01	0	0	0	0	1 *	1 *	1 *
PCF5_01	1	2 *	0	2 *	1 *	1 *	0
RAV1_01	0	1	1	1	2	2	2 *
RAV1_02	0	1	1	1	0	0	1 *
RITA1_01	1	0	0	0	1 *	1 *	0
ROM_Q2	0	2	2	2	1	1	2 *
SBF1_01	0	2 *	2 *	2	1 *	1 *	0
SED_Q2	1	0	0	0	2 *	2 *	1
TAF1_01	0	0	0	0	2 *	2 *	0
TGA1B_01	0	1	1 *	1	0	0	0
TRAB1_Q2	0	3 *	3 *	3 *	1 *	1 *	0
WRKY_Q2	0	2	2	2	2	2	1 *

Table 3. Plant transcription factor binding sites (TFBS) found in the promoter region of genes within the environmentally sensitive (*ms9*, *ms8*, *msp*, and *ms3*) and environmentally stable (*ms2)*soybean male-sterility loci were identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006). Total number of genes in each dataset is in parenthesis below each dataset name in the heading.

Plant TFBS name	***ms9*A[a] (150)**	***ms9*F[e] (123)**	***ms9*SF[i] (35)**	***ms8*A[b] (14)**	***ms8*F[f] (11)**	***ms8*SF[j] (6)**	***msp*A[c] (408)**	***msp*F[g] (337)**	***msp*SF[k] (99)**	***ms3*A[d] (253)**	***ms3*F[h] (206)**	***ms3*SF[l] (86)**	***ms2*A[m] (373)**	***ms2*F[n] (341)**	***ms2*SF[o] (98)**
ABF_Q2 [c,g,k,m,n,o]	52	43	10	4	4	4	149 *	126 *	36 *	87	71	26	108 *	103 *	31 *
ABF1_01	30	25	6	0	0	0	63	57	19	31	27	12	74	72	20
ABF1_02 [a,e,m.n.o]	29 *	25 *	4	1	1	1	70	53	22	38	31	12	59 *	58 *	22 *
ABF1_03 [k]	60	42	11	2	2	2	114	97	33 *	63	58	28	107	106	34
ABI4_01 [k,m,n]	54	42	12	1	1	1	189	156	54 *	127	94	34	181 *	171 *	54
ABZ1_01 [a,c,g,k,m,n,o]	130 *	112	33	8	6	6	374 *	310 *	91 *	254	186	86	329 *	309 *	96 *
AG_01 [a,e,m,n,o]	104 *	91 *	28	7	5	2	205	174	50	184	115	36	202 *	185 *	59 *
AG_02 [c,d,g]	81	75	23	0	0	0	258 *	220 *	53	193 *	134	47	210	200	58
AG_03 [a,e,m,n,o]	104 *	91 *	28	7	5	2	205	174	50	184	115	36	202 *	185 *	59 *
AG_Q2	107	93	22	8	6	4	313	259	66	242	149	64	250	232	57
AGL1_01 [a,e,m,n,o]	124 *	100 *	24	6	5	1	297	249	65	246	151	58	250 *	236 *	74 *
AGL1_02 [a,m,n,o]	116 *	94	22	6	5	1	292	252	64	248	160	60	253 *	241 *	67 *
AGL2_01 [o]	107	91	17	6	4	3	215	186	47	189	115	47	222	202	58 *
AGL2_02	110	95	19	8	6	3	214	179	43	197	120	51	232	212	62
AGL3_01	106	90	17	11	8	4	233	191	46	193	115	50	237	219	59
AGL3_02 [d,h]	97	81	13	12	11	5	258	222	54	215 *	134 *	53	244	225	66

Table 3. (continued)

Plant TFBS name	*ms9*A[a] (150)	*ms9*F[e] (123)	*ms9*SF[i] (35)	*ms8*A[b] (14)	*ms8*F[f] (11)	*ms8*SF[j] (6)	*msp*A[c] (408)	*msp*F[g] (337)	*msp*SF[k] (99)	*ms3*A[d] (253)	*ms3*F[h] (206)	*ms3*SF[l] (86)	*ms2*A[m] (373)	*ms2*F[n] (341)	*ms2*SF[o] (98)
AGL3_03 [d,h]	97	81	13	12	11	5	258	222	54	215 *	134 *	53	244	225	66
AGL15_01 [a,c,d,e,g]	140 *	126 *	32	13	13	8	304 *	262 *	51	244 *	137	53	281	254	66
AGP1_01 [b,j,m,n,o]	71	53	19	13 *	10	7 *	166	141	37	107	69	34	183 *	171 *	50 *
ALFIN1_Q2 [c,g,h,k,n]	376	324	82	30	17	8	1004 *	848 *	286 *	714	497 *	210	835	793 *	229
ANT_01	36	32	10	1	1	0	110	94	31	81	55	23	101	97	16
ARF_Q2[g]	190	164	46	14	6	4	476	403 *	125	386	254	109	469	429	128
ARR10_01	105	86	22	10	7	3	288	234	72	221	146	59	298	271	70
ATHB1_01	68	56	15	2	2	0	169	140	30	172	108	43	161	145	34
ATHB5_01 [o]	125	103	35	12	8	4	342	280	77	281	191	69	348	325	83 *
ATHB9_01	13	11	6	1	1	0	63	53	15	39	22	10	34	30	8
ATMYB15_01 [m,n,o]	13	12	2	2	2	2	34	29	11	34	21	10	41 *	38 *	17 *
ATMYB77_01 [e]	52	46 *	14	4	2	1	110	95	29	111	69	25	126	108	26
ATMYB84_01 [m,n,o]	16	15	1	4	0	0	88	69	20	72	41	18	64 *	61 *	18 *
BHLH66_01 [j,k]	121	99	22	16	11	9 *	408	334	118 *	264	200	79	378	363	111
BPC1_Q2	133	108	27	9	6	4	357	298	81	300	198	82	364	338	96
BZIP910_01 [c,g,k]	27	18	3	0	0	0	68 *	58 *	19 *	58	41	16	65	61	22
BZIP910_02 [o]	39	28	6	3	2	2	104	89	23	45	36	15	102	94	26 *
BZIP911_01 [c,k,o]	25	17	3	2	2	2	62 *	50	17 *	40	24	11	54	51	19 *

Table 3. (continued)

Plant TFBS name	*ms9*A[a] (150)	*ms9*F[e] (123)	*ms9*SF[i] (35)	*ms8*A[b] (14)	*ms8*F[f] (11)	*ms8*SF[j] (6)	*msp*A[c] (408)	*msp*F[g] (337)	*msp*SF[k] (99)	*ms3*A[d] (253)	*ms3*F[h] (206)	*ms3*SF[l] (86)	*ms2*A[m] (373)	*ms2*F[n] (341)	*ms2*SF[o] (98)
BZIP911_02 [a,o]	39 *	27	7	3	3	3	97	86	25	76	53	17	93	90	33 *
C1_Q2 [m,n,o]	140	109	20	15	12	5	450	377	118	322	218	94	462 *	432 *	135 *
CDC5_01	55	48	16	2	2	2	117	97	32	124	66	32	126	122	42
CG1_Q6 [a,c,e,g,k,m,n,o]	200 *	168 *	52	19	14	9	512 *	444 *	153 *	324	248	105	488 *	466 *	154 *
CPRF_Q2 [a,c,e,g,i,k,n,o]	138 *	119 *	40 *	9	7	7	419 *	353 *	136 *	236	184	81	335	312 *	98 *
CPRF1_01 [a,c,e,g,i,k,m,n,o]	82 *	70 *	19 *	7	6	5	263 *	227 *	81 *	148	111	48	207 *	193 *	62 *
CPRF2_01 [a,c,e,g,m,n,o]	79 *	64 *	24	3	3	3	244 *	205 *	56	138	112	48	202 *	194 *	66 *
CPRF2_Q2 [a,c,d,g,m,n,o]	78 *	62 *	22	3	3	3	248 *	210 *	62	139	114	46	212 *	201 *	71 *
CPRF3_01 [a,c,g,i,m,n,o]	81 *	64	24 *	2	1	1	242 *	206 *	58	130	105	48	203 *	194 *	78 *
CPRF3_Q2 [a,c,e,g,i,n,o]	87 *	71 *	25 *	3	2	2	237 *	202 *	59	131	105	46	211	201 *	78 *
DOF_Q2 [c,e,g,h,]	263	215 *	58	27	21	10	776 *	654 *	181	512	357 *	141	629	579	172
DOF1_01 [f]	132	107	22	22	19 *	6	351	296	90	290	184	76	307	290	77
DOF2_01 [g]	155	130	39	18	13	7	379	319 *	89	309	224	103	348	328	97
DOF3_01 [g,n,o]	181	148	34	18	12	4	550	467 *	124	407	284	128	467	439 *	143 *
E2F_Q2 [a,c,e,g,h]	104 *	94 *	28	6	3	1	266 *	224 *	54	167	131 *	52	241	224	51
EMBP1_02 [c,g,i,k]	123	103	29 *	9	8	7	354 *	302 *	98 *	218	160	64	303	279	78
EMBP1_Q2 [a,c,e,g,k,n,o]	105 *	92 *	34	10	7	7	347 *	293 *	99 *	175	139	55	259	242 *	83 *
ERF2_01 [m,n,o]	105	81	26	5	4	2	287	232	66	206	151	53	289 *	272 *	89 *

Table 3. (continued)

Plant TFBS name	*ms9*A[a] (150)	*ms9*F[e] (123)	*ms9*SF[i] (35)	*ms8*A[b] (14)	*ms8*F[f] (11)	*ms8*SF[j] (6)	*msp*A[c] (408)	*msp*F[g] (337)	*msp*SF[k] (99)	*ms3*A[d] (253)	*ms3*F[h] (206)	*ms3*SF[l] (86)	*ms2*A[m] (373)	*ms2*F[n] (341)	*ms2*SF[o] (98)
GAMYB_01 [m]	134	115	35	16	10	5	397	326	94	289	179	85	363 *	318	102
GAMYB_Q2 [g,k]	168	138	37	20	12	8	489	414 *	128 *	381	257	120	461	428	117
GBF_Q2 [c,g,k,m,n,o]	113	91	37	9	7	7	336 *	284 *	96 *	199	159	68	293 *	280 *	92 *
GBP_Q6 [a,c,g,k,m,n,o]	154 *	124	48	12	10	10	467 *	387 *	120 *	273	216	89	414 *	400 *	160 *
GT1_Q6 [g,m,n,o]	63	53	13	3	3	2	160	140 *	36	136	93	38	170 *	155 *	53 *
HAHB4_01 [h,i,k]	95	81	26 *	5	0	0	290	240	83 *	230	172 *	65	261	241	69
HBP1A_Q2 [a,c,e,g,k,o]	139 *	118 *	45	9	6	6	371 *	316 *	102 *	200	168	72	316	300	117 *
HBP1B_Q6 [h,n,o]	100	78	27	5	4	3	331	278	71	202	141 *	55	307	286 *	92 *
HBPA1_Q6_01 [c,g,n,o]	116	96	34	3	1	1	372 *	314 *	84	254	181	79	345	324 *	103 *
ID1_01 [g]	188	150	40	17	12	7	516	440 *	129	363	245	104	433	396	117
KNOX3_01 [c,e,k,m,n,o]	175	146 *	46	8	7	2	456 *	399 *	129 *	308	220	95	390 *	359 *	107 *
LIM1_01 [c,g,k,l,m,n]	269	228	52	25	13	4	837 *	690 *	225 *	612	433	184 *	718 *	672 *	205
MADSA_Q2 [e]	160	140 *	36	14	11	7	456	377	108	335	214	82	411	377	99
MADSB_Q2	150	128	28	13	11	5	398	336	90	296	180	76	339	314	80
MYB80_01	48	40	7	3	2	0	120	79	31	92	62	20	128	121	27
MYBAS1_01 [g,k,m,n]	146	123	28	20	17	5	435	362 *	103 *	316	219	98	399 *	356 *	95
MYBPH3_01	32	28	8	3	3	2	72	58	16	55	39	11	73	65	23
MYBPH3_02 [d]	39	31	11	4	4	3	175	140	46	112 *	69	34	119	104	32

Table 3. (continued)

Plant TFBS name	*ms9*A[a] (150)	*ms9*F[e] (123)	*ms9*SF[i] (35)	*ms8*A[b] (14)	*ms8*F[f] (11)	*ms8*SF[j] (6)	*msp*A[c] (408)	*msp*F[g] (337)	*msp*SF[k] (99)	*ms3*A[d] (253)	*ms3*F[h] (206)	*ms3*SF[l] (86)	*ms2*A[m] (373)	*ms2*F[n] (341)	*ms2*SF[o] (98)
NAC691_01 [c,d,g,h,k,l]	13	10	0	0	0	0	63 *	51 *	11 *	56 *	44 *	26 *	62	61	23
O2_01	32	24	4	4	4	2	87	74	27	66	44	21	94	86	18
O2_02 [a,c,e,g,k,m,n,o]	117 *	93 *	35	6	6	4	382 *	319 *	95 *	210	166	71	280 *	265 *	98 *
O2_03	126	111	25	15	10	5	369	304	95	314	199	81	394	364	96
O2_Q2 [c,g,m,n,o]	93	74	23	12	10	5	306 *	259 *	66	205	136	53	292 *	279 *	76 *
OCSBF1_01 [c,g,i,k,m,n,o]	169	139	45*	13	10	8	513 *	430 *	124 *	299	198	85	434 *	398 *	126 *
OSBZ8_Q6 [c,g,k,m,n,o]	137	109	33	12	10	7	475 *	412 *	135 *	286	219	97	386 *	371 *	115 *
P_01 [l]	131	101	22	12	9	4	401	329	110	309	211	91 *	333	311	100
PBF_01 [c,e,g]	195	163 *	52	18	13	6	476 *	398 *	117	369	245	108	391	366	104
PBF_Q2	210	174	38	12	10	4	547	469	133	494	317	140	484	444	114
PCF2_01 [c,d,g,h,k,l,m,n]	138	118	29	10	7	5	398 *	333 *	126 *	282 *	195 *	94 *	354 *	325 *	94
PCF5_01 [c,g,k,m,n,o]	127	106	17	7	4	4	300 *	262 *	89 *	200	124	60	271 *	255 *	75 *
PEND_01 [o]	89	68	20	8	5	1	291	240	59	231	145	53	238	218	58 *
PIF3_01 [e,n,o]	53	44 *	9	5	4	4	167	136	48	100	81	33	134	131 *	42 *
PIF3_02 [a,c,e,g,m,n,o]	51 *	46 *	6	4	4	4	147 *	119 *	40	82	65	32	131 *	128 *	32 *
RAV1_01 [d,h,l,m,n]	198	165	47	24	17	7	529	429	141	463 *	303 *	132 *	533 *	492 *	138
RAV1_02 [a,m,n,o]	123 *	94	24	12	8	5	301	253	73	223	137	50	351 *	321 *	95 *
RITA1_01 [c,g,i,k,m,n,o]	127	102	39 *	9	8	7	443 *	371 *	99 *	259	187	85	363 *	341 *	117 *

Table 3. (continued)

Plant TFBS name	*ms9*A[a] (150)	*ms9*F[e] (123)	*ms9*SF[i] (35)	*ms8*A[b] (14)	*ms8*F[f] (11)	*ms8*SF[j] (6)	*msp*A[c] (408)	*msp*F[g] (337)	*msp*SF[k] (99)	*ms3*A[d] (253)	*ms3*F[h] (206)	*ms3*SF[l] (86)	*ms2*A[m] (373)	*ms2*F[n] (341)	*ms2*SF[o] (98)
ROM_Q2 [m,n,o]	163	140	42	18	12	5	470	390	105	383	250	99	455 *	422 *	120 *
SBF1_01 [c,g]	67	56	17	2	1	1	232 *	193 *	47	159	104	46	210	198	59
SED_Q2 [c,g,k]	180	151	37	22	15	5	497 *	429 *	120 *	371	240	98	435	396	113
SPF1_Q2	69	47	12	5	3	1	242	205	61	192	125	46	232	212	67
TAF1_01 [a,c,d,g,k,m,n,o]	196 *	162 *	54	8	5	4	558 *	469 *	141 *	340	258	118	505 *	477 *	167 *
TAF1_Q2 [a,c,e,g,k,n,o]	154 *	125 *	49	10	8	8	437 *	367 *	117 *	252	192	88	383	363 *	132 *
TEIL_01 [m]	93	78	18	7	5	2	283	230	77	197	118	49	263 *	232	55
TGA1A_Q2 [m,n,o]	143	109	46	8	3	3	480 *	408 *	116	248	189	77	388 *	371 *	136 *
TGA1A_Q2_01	87	63	21	10	6	1	272	236	61	242	164	72 *	256	239	73 *
TGA1B_01 [a,g,m,n,o]	91 *	69	24	5	4	3	235	200 *	50	146	106	43	217 *	209 *	74 *
TGA1B_Q2 [a,c,g,m,n,o]	154 *	123	39	7	5	5	421 *	348 *	85	271	202	92	409 *	391 *	150 *
TRAB1_Q2 [a,c,e,g,k,m,n,o]	184 *	148 *	32	13	8	7	490 *	420 *	153 *	322	247	113	458 *	441 *	143 *
WRKY_Q2 [m,n]	121	100	26	3	3	1	319	271	69	241	163	57	294 *	266 *	71
ZAP1_01 [n,o]	70	64	12	4	4	3	178	141	40	117	84	38	162	156 *	44 *
GRAND TOTAL	11395	9425	2621	910	664	390	31986	26873	7991	22709	15669	5980	28822	26893	8183

* Significantly overrepresented TFBS, $t \leq 0.05$.

[a,b,c,d] All genes in environmentally sensitive male-sterility loci (*ms9*, *ms8*, *msp*, and *ms3*) respectively.

[e,f,g,h] Genes with at least 1 floral read in SoySeq (Severin et al., 2010) for environmentally sensitive male-sterility loci (*ms9*, *ms8*, *msp*, and *ms3*) respectively.

Table 3. (continued)

[i,j,k,l] Genes with floral expression significantly overrepresented determined using Fisher's exact test (Fisher, 1949) with Bonferroni correction (P-value ≤ 0.05, Bonferroni, 1936) for environmentally sensitive male-sterility loci (*ms9*, *ms8*, *msp*, and *ms3*) respectively.

[m,n,o] Non-environmentally sensitive male-sterility locus, *ms2*, all genes, < 1 floral read, significantly floral datasets, respectively.

Table 4. Homeologous regions of molecularly mapped male-sterility loci in soybean.

Male-sterility homeologous region	All tissues[†]	Floral[†]	Predicted genes	Predicted genes with > 1 floral reads	Predicted genes with significantly floral expression[‡]	Chromosome number and location (bp)	Region size (Kb)	MLG	Gene range from Glyma1 genome assembly[§]
*ms2*H	33,251	1,164,950	510	403	110	Gm20:33040389..36501513 ; 45993589..46721826	3461.1; 728.2	I	Glyma20g17310 to 20g23090; 20g38380 to 20g39410
*ms8*H	872	7,884	20	13	7	Gm08:18310741..18484951	175.7	A2	Glyma08g24080 to 08g24290
*ms9*H	7,061	113,055	153	134	40	Gm19:41176292..42986270	1810.0	L	Glyma19g33590 to 19g35420
*ms3*H_*Gm01*	11,477	131,813	375	171	N/A[¶]	Gm01:5005455..6081494; 6314284..7786604; 8045243..9175815; 9446877..9645158; 9915227..10036502; 12744692..13070514; 27963671..28780308; 29393584..29846090	1076.0; 1472.3; 1130.6; 198.2; 121.2; 325.8; 816.6; 452.5	D1a	N/A[#]
*ms3*H_*Gm07*	2,324	39,819	23	15	N/A[¶]	Gm07:38607618..38835099	227.5	M	Glyma07g33640 to 07g33890
*msp*H_Gm14	17,436	225,465	414	314	N/A[¶]	Gm14:44224418..48793396	4569.0	B2	Glyma14g35330 to 14g39710
*msp*H_Gm10	1,310	11,572	42	22	N/A[¶]	Gm10:9707589..10305601	589.0	O	Glyma10g10110 to 10g10560
*msp*H_Gm17	4,191	66,877	66	56	N/A[¶]	Gm17:5707946..6200215	492.2	D2	Glyma17g07700 to 17g08360

Table 4. (continued)

† Raw expression counts from 14 tissue types including root, nodule, young leaf, flower, pod, and seed with different time points during pod and seed development, RNA-Seq atlas (Severin et al., 2010).

‡ Significantly floral genes identified by Fisher's exact test (Fisher, 1949) comparing read distribution for each gene relative to the soybean genome. A Bonferroni correction was applied for over-sampling (P-value ≤ 0.05, Bonferroni, 1936).

§Glyma 1, Schmutz et al. (2010) and SoyBase (www.soybase.org).

¶ Fisher's exact test not performed because partial homeologs were not analyzed further.

Details in Supplemental text file.

Table 5. Plant transcription factor binding sites (TFBS) identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006) for three complete homeologs (*ms9*H, *ms8*H, and *ms2*H) to the male-sterility loci *ms9*, *ms8*, and *ms2* in soybean. Total number of genes in each dataset is in parenthesis below each dataset name in the heading.

Plant TFBS Name	***ms9*HA**[a] **(153)**	***ms9*HF**[c] **(143)**	***ms9*HSF**[e] **(40)**	***ms8*HA**[b] **(20)**	***ms8*HF**[d] **(13)**	***ms8*HSF**[f] **(7)**	***ms2*HA**[g] **(510)**	***ms2*HF**[h] **(403)**	**ms2HSF**[i] **(110)**
ABF_Q2[a,c,e]	50 *	41 *	17 *	4	4	2	157	132	20
ABF1_01	31	25	11	3	2	1	79	70	19
ABF1_02	21	19	6	2	1	0	74	63	9
ABF1_03	45	38	3	8	5	4	120	101	15
ABI4_01[c,g,h,i]	82	77 *	26	12	8	5	243 *	218 *	54 *
ABZ1_01[g,h]	143	128	48	15	8	6	428 *	372 *	85
AG_01	75	67	18	7	6	4	256	215	57
AG_02[a,c]	88 *	77 *	22	11	7	3	277	236	56
AG_03	75	67	18	7	6	4	256	215	57
AG_Q2	120	108	36	6	3	1	337	286	71
AGL1_01[h,i]	92	83	21	11	9	5	350	290 *	75 *
AGL1_02	82	74	18	11	9	6	372	315	82
AGL2_01	79	66	15	13	12	5	261	209	63
AGL2_02	83	69	18	10	10	4	270	221	64
AGL3_01	82	73	18	13	12	7	262	211	61
AGL3_02	82	68	16	12	10	5	283	234	73
AGL3_03	82	68	16	12	10	5	283	234	73
AGL15_01	115	95	20	12	12	6	368	305	77
AGP1_01	70	53	18	17	12	4	245	198	46
ALFIN1_Q2[a,c, g,h]	349 *	309 *	98	42	30	19	1157 *	983 *	217
ANT_01	37	34	15	4	2	1	117	101	23
ARF_Q2	196	169	60	21	13	7	637	531	136
ARR10_01[d]	114	94	29	21	14 *	5	372	309	84
ATHB1_01	68	62	19	7	6	4	231	176	43
ATHB5_01	121	104	35	17	13	7	432	337	99

Table 5. (continued)

Plant TFBS Name	*ms9*HA[a] (153)	*ms9*HF[c] (143)	*ms9*HSF[e] (40)	*ms8*HA[b] (20)	*ms8*HF[d] (13)	*ms8*HSF[f] (7)	*ms2*HA[g] (510)	*ms2*HF[h] (403)	ms2HSF[i] (110)
ATHB9_01	9	9	2	0	0	0	47	40	8
ATMYB15_01[g,h]	22	22	7	1	0	0	61 *	49 *	6
ATMYB77_01[a,c]	56 *	51 *	14	8	2	1	153	122	26
ATMYB84_01	25	21	8	2	2	0	69	63	12
BHLH66_01[f,h]	159	146	34	20	13	9 *	467	386 *	85
BPC1_Q2[c]	144	131 *	39	26	18	11	479	388	97
BZIP910_01	22	17	10	1	1	0	78	65	12
BZIP910_02	31	27	6	2	0	0	117	94	24
BZIP911_01	14	11	7	1	1	0	86	71	13
BZIP911_02[e]	36	29	12 *	5	2	1	141	117	31
C1_Q2[b,d,f,g,h,i]	153	139	45	30 *	20 *	11 *	580 *	478 *	147 *
CDC5_01	40	37	10	11	7	0	188	153	41
CG1_Q6[a,c,g,h]	195 *	176 *	48	22	12	5	651 *	551 *	136
CPRF_Q2[a,c,e,g,h]	133 *	116 *	40 *	8	6	2	494 *	409 *	98
CPRF1_01[a,c,e,g,h]	90 *	77 *	23 *	8	5	3	304 *	254 *	61
CPRF2_01[g,h]	84	72	22	8	4	4	270 *	219 *	53
CPRF2_Q2[a,c,e,g,h]	87 *	74 *	25 *	8	5	4	282 *	233 *	56
CPRF3_01[a,c,e,g,h]	85 *	75 *	23 *	10	7	5	283 *	235 *	56
CPRF3_Q2[a,c,e,h]	84 *	73 *	22 *	11	8	5	280	232 *	55
DOF_Q2[a,c,e]	279 *	253 *	87 *	26	21	12	754	631	144
DOF1_01	146	135	44	26	15	9	430	355	80
DOF2_01[b]	142	131	43	29 *	18	8	461	385	102
DOF3_01[b,d,f]	188	167	51	32 *	22 *	13 *	662	556	141
E2F_Q2[a,c,h]	105 *	89 *	19	11	6	3	309	243 *	64
EMBP1_02[e,g,h]	120	105	36 *	10	8	6	388 *	314 *	72
EMBP1_Q2[a,c,e,g,h]	100 *	82 *	35 *	14	11	6	364 *	298 *	66
ERF2_01	118	100	24	15	7	3	335	276	67
GAMYB_01[d,f,i]	141	126	31	21	16 *	7 *	505	410	122 *

Table 5. (continued)

Plant TFBS Name	*ms9*HA[a] (153)	*ms9*HF[c] (143)	*ms9*HSF[e] (40)	*ms8*HA[b] (20)	*ms8*HF[d] (13)	*ms8*HSF[f] (7)	*ms2*HA[g] (510)	*ms2*HF[h] (403)	ms2HSF[i] (110)
GAMYB_Q2	177	157	42	25	9	4	596	476	132
GBF_Q2[e,g,h]	116	105	36 *	13	8	4	410 *	344 *	76
GBP_Q6[e,g,h]	170	150	58 *	17	10	5	534 *	450 *	111
GT1_Q6	58	50	10	11	8	7	187	147	37
HAHB4_01[a,c,f]	121 *	107 *	29	12	8	6 *	358	285	74
HBP1A_Q2[a,c,e,h]	150 *	130 *	44 *	16	13	9	457	389 *	94
HBP1B_Q6[g,h]	114	98	29	14	5	1	389 *	332 *	74
HBPA1_Q6_01[g,h]	133	116	37	16	8	4	491 *	428 *	87
ID1_01	190	173	62	33	22	11	522	424	120
KNOX3_01[a,c,g,h]	172 *	151 *	39	19	10	6	546 *	447 *	107
LIM1_01[b,d,f,g,h,i]	282	238	67	50 *	36 *	23 *	938 *	789 *	219 *
MADSA_Q2	166	146	34	20	11	7	545	446	114
MADSB_Q2	138	121	22	25	14	7	482	394	105
MYB80_01	41	39	8	2	1	0	171	137	41
MYBAS1_01[d,f,g,h]	141	122	35	28	18 *	10 *	522 *	438 *	117
MYBPH3_01	27	23	6	8	3	2	129	101	28
MYBPH3_02[h]	53	46	11	6	4	2	190	147 *	45
NAC691_01	24	20	3	3	1	0	74	66	21
O2_01	27	24	8	0	0	0	110	87	34
O2_02[h]	127	109	40	15	9	9	438	358 *	80
O2_03	130	105	42	22	14	12	461	370	97
O2_Q2	75	70	12	12	5	5	364	302	81
OCSBF1_01[a,c,g,h,i]	191 *	169 *	52	20	16	12	593 *	490 *	130 *
OSBZ8_Q6	149	134	38	18	8	7	445	382	94
P_01[g,h]	160	137	44	18	12	5	537 *	434 *	128
PBF_01	199	181	60	25	15	9	558	467	114
PBF_Q2	211	189	58	28	19	9	722	613	158
PCF2_01[h]	118	107	45	12	9	5	499	427 *	107

Table 5. (continued)

Plant TFBS Name	*ms9*HA[a] (153)	*ms9*HF[c] (143)	*ms9*HSF[e] (40)	*ms8*HA[b] (20)	*ms8*HF[d] (13)	*ms8*HSF[f] (7)	*ms2*HA[g] (510)	*ms2*HF[h] (403)	ms2HSF[i] (110)
PCF5_01[h]	82	71	26	9	4	1	315	276 *	76
PEND_01	103	87	22	16	12	5	397	304	77
PIF3_01	54	47	18	5	5	2	199	166	37
PIF3_02[e,g,h]	47	41	19 *	2	1	1	181 *	150 *	24
RAV1_01[g,h]	189	168	52	24	16	8	694 *	533 *	142
RAV1_02[b]	141	126	31	25 *	11	3	386	314	74
RITA1_01[g,h]	145	124	39	16	10	8	499 *	398 *	100
ROM_Q2[g,h]	168	143	56	27	14	8	607 *	516 *	145
SBF1_01	73	71	18	8	8	3	259	209	57
SED_Q2[e]	179	163	65 *	19	11	7	589	493	134
SPF1_Q2	92	75	28	6	5	2	317	250	68
TAF1_01[a,c,g,h]	218 *	188 *	54	25	16	11	648 *	540 *	123
TAF1_Q2[a,c,e,h]	158 *	137 *	43 *	18	12	8	507	420 *	88
TEIL_01[f]	93	88	25	18	13	8 *	349	275	84
TGA1A_Q2[g,h]	163	140	42	20	12	11	536 *	443 *	99
TGA1A_Q2_01[g,h]	84	72	25	17	11	4	368 *	302 *	79
TGA1B_01[a,c,h]	91 *	79 *	25	8	7	4	278	232 *	54
TGA1B_Q2[a,g,h]	167 *	146	46	14	6	2	547 *	462 *	100
TRAB1_Q2[g,h]	178	159	40	20	13	10	563 *	488 *	105
WRKY_Q2	133	110	29	16	12	5	391	311	84
ZAP1_01	52	48	11	6	4	3	225	172	51
GRAND TOTAL	11460	10059	3073	1471	960	543	38228	31613	7960

* Significantly overrepresented TFBS, $t \leq 0.05$.

[a,b,c,d] All genes in environmentally sensitive male-sterility homeologs, *ms9* and *ms8* respectively.

[e,f,g,h] Genes with at least 1 floral read in SoySeq (Severin et al., 2010) for environmentally sensitive male-sterility homeologs, *ms9* and *ms8* respectively.

[i,j,k,l] Genes with floral expression significantly overrepresented determined using Fisher's exact test (Fisher, 1949) with Bonferroni correction (P-value ≤ 0.05, Bonferroni, 1936) for environmentally sensitive male-sterility homeologs, *ms9* and *ms8* respectively.

Table 5. (continued)

[m,n,o] Non-environmentally sensitive male-sterility homeolog, *ms2*H, all genes, < 1 floral read, significantly floral datasets, respectively.

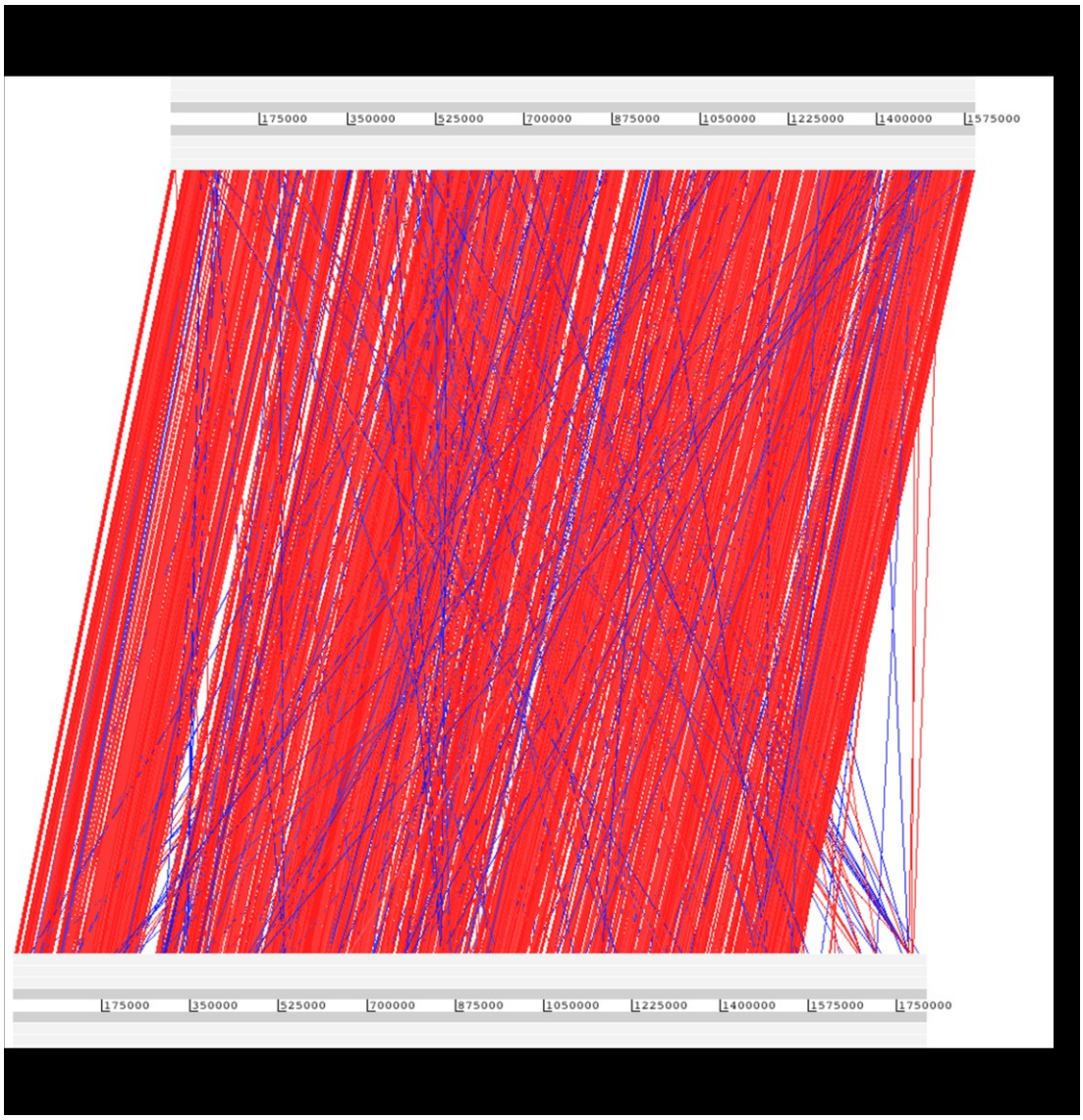

Figure 1. BLASTN ($E < 10^{-4}$, Altschul et al., 1997) of the *ms9* locus to the homeologous region on Gm19 visualized by Artemis comparison tool (ACT, Carver et al., 2005).

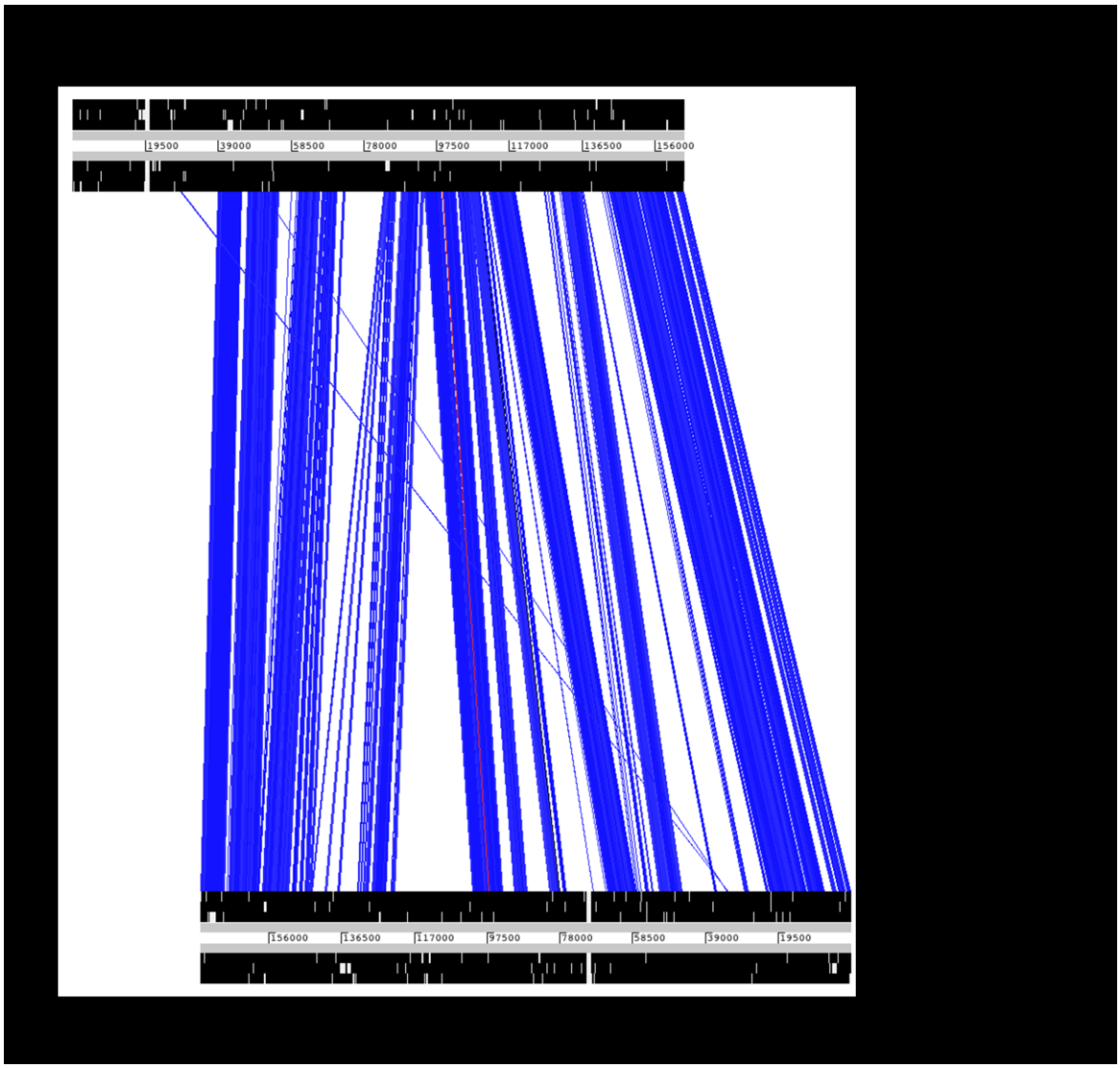

Figure 2. BLASTN ($E < 10^{-4}$, Altschul et al., 1997) of the *ms8* locus to homeologous region on Gm08. Using the Artemis comparison tool (ACT, Carver et al., 2005) the homeologous region has been flipped to reverse orientation to clearly visualize similarity between *ms8* and *ms8*H.

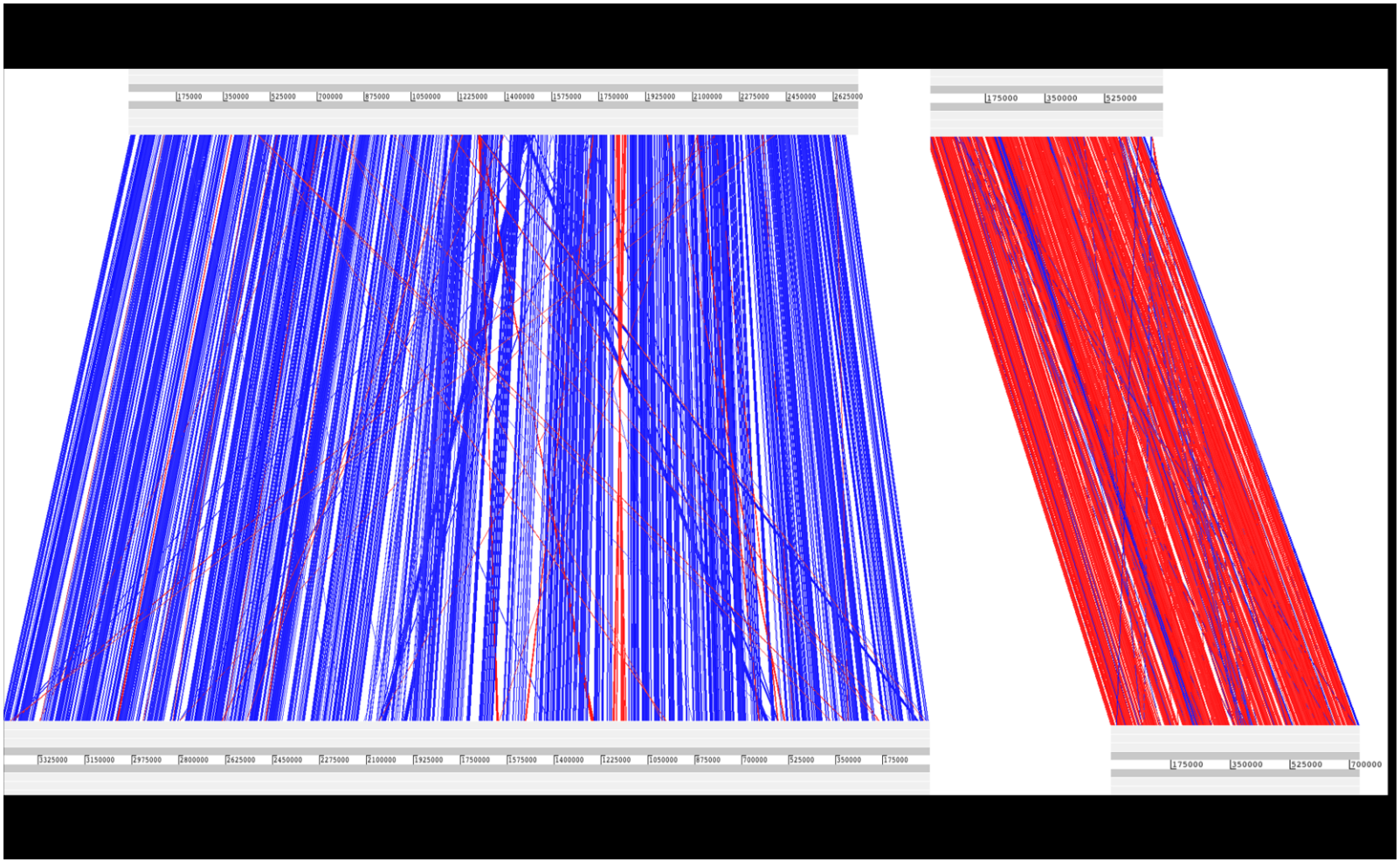

Figure 3. BLASTN ($E < 10^{-4}$, Altschul et al., 1997) of the *ms2* locus to the homeologous regions on Gm10, the 9.49 Mb nonhomologous segment has been removed from *ms2*H and the two fragments were treated as a complete homeolog. For *ms2*H, the first fragment has been flipped to reverse orientation for clear view of similarity by Artemis comparison tool (ACT, Carver et al., 2005).

CHAPTER 4: IDENTIFICATION AND EXPRESSION ANALYSES OF CANDIDATE GENES FOR MALE-STERILITY LOCUS *ms9* IN SOYBEAN

A paper for submission to Theoretical and Applied Genetics

Christiana E. Wiebbecke[1,2], Silvia R. Cianzio[2], Reid G. Palmer[2,3], and Michelle A. Graham[2,3,4]

ABSTRACT

The increased growth rate of the human population has driven the need to increase soybean yield at a faster pace than traditional breeding methods achieve. The *ms9* locus has greater insect pollinator attraction and out-crossed seed-set than other male-sterility loci in soybean enhancing its potential for use in hybrid seed production. However, day temperature influences the male-sterile phenotype, which makes it difficult to identify male-sterile, female-fertile plants prior to flowering for use as the female (pod) parent. The objectives of this research were to use mapping information and available genomic resources to identify candidate genes for *ms9* and to analyze gene expression patterns of fertile and male-sterile, female-fertile plants during floral development across controlled environments. By combining bioinformatic methods and review of the available thermosensitive genic-male sterility literature, we identified nine candidate genes for *ms9*. Quantitative real-time polymerase chain reaction (qRT-PCR) was used to measure candidate gene expression. Two genes, Glyma03g30880 and Glyma03g31530, had increased expression in both fertile and

[1] Monsanto Co., 800 North Lindbergh Boulevard, St. Louis, MO 63167

[2] Dep. of Agronomy, Iowa State Univ., Ames, IA 50011

[3] USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA 50011. Joint collaboration of Monsanto, the Iowa Agric. and Home Econ. Exp. Stn., Ames, Iowa, Project No. 4403, and the USDA-ARS, Corn Insects and Crop Genet. Res. Unit., Ames, IA.

[4] Corresponding author (Michelle.Graham@ARS.USDA.GOV)

male-sterile floral samples as day temperature increased from 30 to 35 °C. Sequence homology to known male-sterility genes and temperature-dependent expression make Glyma03g30880 and Glyma03g31530 the most likely candidates for *ms9*.

INTRODUCTION

The world human population is predicted to exceed 9 billion by 2050 (Population Reference Bureau 2011). To meet demand, food production must increase by over 70% (Food and Agriculture Organization 2009). Soybean is the largest source of protein feed and the second largest source of vegetable oil in the world (Economic Research Service 2010). While human population dynamics demand an increase in agricultural production, biofuels will directly compete with foodstocks and feedstocksfor limited land and resources. Improving yield while minimizing the impacts of pathogens and pests is an essential component of this challenge.

Soybean is a highly autogamous crop in which traditional breeding methodology has focused on the selection of elite lines within segregating families developed by manual cross-pollination. This approach has achieved an average yield increase of 15.1 $kg \cdot ha^{-1}$, or 0.6%, per year since the 1920s (Specht and Williams 1984). However, there is a permanent economic and food production need to increase yield at even faster pace. As breeders look for new methods to increase yield, development of hybrid cultivars may be an alternative approach to increase soybean yield. Hybrids in other self-pollinated crops have shown heterosis, and in some cases superior performance compared to parental lines. Hybrid yield increases in other economically important self-pollinated crops ranged from 15 to 45% in

rice (Li and Yuan 2000; Virmani et al. 2003) and from 14 to 64% in pigeonpea (Saxena et al. 2010). Similarly, cotton hybrids have greater yield and flint quality traits compared to mid-parent values (McCarty et al. 2004a, 2004b; Jenkins et al. 2009). Therefore, hybrid production could become a practical approach for improving yield in soybean.

In self-pollinated crops, the amount of out-crossed seed produced is directly related to the degree of insect pollinator attraction, which ties directly to potential hybrid yield and yield stability (Palmer et al. 2009). In autogamous crops, male sterility can be used to reduce self-pollinated seed-set and promote hybrid seed production. To date, 11 male-sterility loci (*ms1* thru *ms9* and *msp*, Palmer et al. 2001) and *ms MOS* (Midwest Oil Seeds, Jin et al. 1997) have been identified in soybean. Palmer (2000) determined the *ms9* male-sterility locus is a single recessive nuclear gene, which is expected to segregate in a phenotypic ratio of three fertile plants: one male-sterile, female-fertile plant. Further evaluation conducted by Ortiz-Perez et al. (2006) determined the seed-set achieved by insect-mediated cross-pollination was highest for *ms9* lines, indicating a possible high insect pollinator attraction to plants carrying this gene. These results suggest that *ms9* could be a valuable tool in soybean hybrid seed production.

A critical challenge in implementing male-sterile systems for soybean variety development is the reliable identification of male-sterile, female-fertile plants prior to flowering (R2 stage, Fehr and Caviness 1977). Classification of fertile versus male-sterile, female-fertile plants is traditionally accomplished by examining the pollen phenotype at flowering, which causes delays in cross-pollination and is labor intensive. An alternative to pollen phenotypic classification is to develop a molecular marker that can be used in marker assisted selection (MAS) programs to identify male-sterile, female-fertile plants early in plant

development. This would enable breeders to select male-sterile, female-fertile plants for cross-pollination leading to a more efficient hybridization system.

In the case of *ms9*, Wiebbecke et al. (2011a) determined that there were no obvious differences in pollen shed at anthesis that could be used to distinguish fertile from male-sterile, female-fertile plants, at R2 stage. Further investigation demonstrated the *ms9* locus was influenced by day temperature, thus identification and maintenance of male-sterile phenotype was dependent on day temperature. Male-sterile, female-fertile plants revert to male fertility at low day temperatures (30 °C) and maintain male sterility at high day temperatures (35 °C) (Wiebbecke et al. 2011a). This thermosensitive genic male-sterile (TGMS) system could be readily exploited with a more efficient, reliable method to identify male-sterile, female-fertile plants prior to flowering.

At the molecular level, the *ms9* locus was mapped between simple sequence repeat (SSR) markers Satt521 and Satt237 on MLG N, and linked to each marker by 8.5 centi-morgan (cM) and 16.2cM, respectively (Cervantes-Martinez et al. 2007). Our prior bioinformatic analyses identified homologs of previously identified Arabidopsis and rice male-sterility genes within the mapped soybean male-sterility loci *msp*, *ms3*, and *ms2* (Wiebbecke et al. 2011b). No male-sterility homologs, however, were identified for the *ms9* locus. Therefore, objectives of our research were to use the mapping information (Cervantes-Martinez et al. 2007) and available genomic resources to identify candidate male-sterility genes for *ms9* and analyze gene expression patterns of candidate genes among fertile and male-sterile, female-fertile plants during floral development across controlled environments.

MATERIALS AND METHODS

Plant Material

Soybean genetic type number T359H, the heterozygous maintainer line for *ms9*, was generated from the initial cross of T325 × L67-3483 (Palmer 2000). Self-pollination of progeny from T359H (*Ms9ms9*) plants are expected to generate plants with the approximate phenotypic ratio 3 fertile (*Ms9Ms9* or *Ms9ms9*) plants: 1 male-sterile, female-fertile (*ms9ms9*) plant.

Identification of Candidate Genes

ms9 was mapped between simple sequence repeat (SSR) markers Satt521 and Satt237 on MLG N (Cervantes-Martinez et al. 2007). The flanking SSR markers were used to query the soybean genome browser (www.soybase.org) to visualize the *ms9* genomic region, on chromosome Gm03. Every gene in the region was examined for involvement in male sterility using several steps. First, gene annotation information provided by SoyBase (www.soybase.org), was used in conjunction with the words thermosensitive genic male-sterility to search the literature for involvement in male sterility, influenced by temperature. Second, the predicted coding sequences were searched using BLASTX (Altschul et al. 1997) against the non-redundant protein sequences in GenBank (blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm gene annotation. The GenBank accession with the most informative putative annotation and highest homology was obtained (Table 1). If the gene could be linked to male sterility or floral development, it became a potential candidate. We used two methods to determine if the candidate gene was expressed in floral tissue. The first method used BLASTN ($E<10^{-6}$, Altschul et al. 1997) to search the predicted coding sequence of candidate genes against the soybean expressed sequence tags (EST) (Dana Faber Cancer Institute Gene

Indices, Lee et al. 2005, compbio.dfci.harvard.edu/tgi/tgipage.html) to identify genes with floral EST library identifiers (Gm-c1015, c1016, c1051, and c1061, Shoemaker et al. 2002). Second, we used the recently available RNA-Seq atlas (Severin et al. 2010, www.soybase.org/soyseq/) to determine the expression of potential candidate genes in floral tissue. For every gene, we recorded the number of raw (unnormalized) floral and nonfloral tissue reads. Fisher's exact test was used to identify those genes significantly overrepresented with floral reads by comparing read distribution in each gene relative to all genes in the genome (Fisher 1949). To correct for oversampling, a Bonferoni correction was applied (P-value ≤ 0.05, Bonferroni 1936). Finally, we used TBLASTX ($E<10^{-6}$, Altschul et al. 1997) to compare candidate gene coding sequences to the predicted proteins in The Arabidopsis Information Resource (TAIR, www.arabidopsis.org). Once an Arabidopsis homolog was identified, we used expression information, specifically noting floral tissues, and phenotype information from Arabidopsis mutants to prioritize further candidate genes. This search methodology identified nine genes, classified into seven functional groups, as possible candidates for *ms9* (Table 1).

Floral Sample Collection

In total, 448 selfed progeny of T359H were grown in seven environments. 56 plants were grown in growth chambers with one of the following night/ day temperature regimes: 12/30, 15/30, 18/30, 23/30, and 18/35 °C with day length 13 h at flowering, 21/30 and 17/35 °C with 14 h day length at flowering, and temperature regime 23/30 °C was repeated once (Wiebbecke et al. 2011a). At approximately six weeks, corresponding to the R2 stage, reproductive buds and open flowers were evaluated for pollen phenotype visually and by microscopy to confirm fertile or male-sterile pollen phenotype (Wiebbecke et al., 2011a).

Concurrently, we collected all available reproductive buds approximately one day before anthesis and at anthesis on each plant in each environment. Floral samples were flash frozen in liquid nitrogen and stored at -80 °C. We attempted to equally represent male-sterile, female-fertile and fertile floral samples, because of expected segregation of the recessive *ms9* locus. All (102) male-sterile floral samples were submitted for candidate gene expression and we selected a subset of fertile (118) samples encompassing the range of fertile selfed seed-set in each of the seven environments.

Floral RNA Extraction and Quantitative RT-PCR

In total, 220 floral samples were submitted for floral expression analysis (102 male-sterile and 118 fertile samples) from the seven environments described above. Floral samples were not pooled across fertile or male-sterile floral phenotypic classes nor were they pooled across environments, to have more power to detect subtle differences in fertile and/or male-sterility gene expression across environments. RNA was extracted from floral tissue using the E-Z 96 Plant RNA Purification Kit (Omega BioTek, R1027-02). Eight samples were discarded due to insufficient RNA yield or quality (6 male-sterile and 2 fertile samples from the 18/35 °C environment). RNA samples were treated with Turbo DNA-Free DNase (Ambion, AM1907) and quantified by spectrophotometery. Dilutions based on the concentration were made to normalize the samples to 5 ng/μL final concentration. Samples were set up in triplicate blocks to serve as technical replicates during the qRT-PCR analysis.

Primers/probe sets for qRT-PCR were selected using Applied Biosystems Primer Express version 2.0 software. Initial probe and primer design parameters were melting temperature (Tm) 61 to 63 °C, with 62 °C optimum; GC content from 40 to 60%, and amplicon length between 50 and 150 base pairs. We relaxed the Tm parameter because of

required specificity for candidate gene targets (Supplemental Table 1). All assays were validated using standard curve validation procedures targeting an amplification efficiency of 90 to 110% and slope of -3.3 ± 0.3 (Applied Biosystems 2008). For validation, serial dilutions ranged from 250 to 0.007 ng/µL of target sequence of soybean genomic DNA. qRT-PCR was carried out using the TaqMan® One-Step RT-PCR Master Mix Reagents Kit (Applied Biosystems, 4309169) in an ABI7900HT following Applied Biosystems manufacture recommendations with a 60 °C annealing temperature and 40 cycles. A final primer concentration of 300 nM and 200 nM probe was used for each reaction.

Identification of Candidate Gene Homeologs and Assay Specificity

To identify homeologs for each of the nine candidate male-sterility genes above, we first downloaded the sequence for each candidate gene using the SoyBase genome browser (www.soybase.org). We then used BLASTN ($E < 10^{-20}$, Altschul et al. 1997) to identify homeologs for the male-sterility candidate genes (Table 1). Identified homeologs were used in primer and probe design to ensure specificity of candidate gene primer and TaqMan® probe sets. Custom DNA Oligo templates corresponding to each homeolog that had potential to amplify were ordered ((250nmole DNA oligo standard desalting, Integrated DNA Technologies), Supplemental Table 1). Homeolog templates were diluted based on the molecular weight of each amplicon to a molar ratio of 1 copy amplicon in 50 ng/µL corn genomic DNA, which served as a carrier molecule (Pfaffl 2002). During assay validation, described above, homeolog templates were tested to confirm no amplification was detected (40 cycle thresholds, Ct).

Statistical Analysis

The internal control gene actin (Glyma19g00850, an endogenous housekeeping gene) was used to normalize the amount of messenger RNA (mRNA) present in the total RNA of each sample. First, the mean Ct for each sample was calculated from the triplicate technical replicates. Then, it was necessary to calculate the mean relative quantity of actin for each sample, using y = 10^(mean sample Ct – b)/m), where b = slope and m = intercept from the standard curve validation (Applied Biosystems 2001) (Supplemental Table 2). After determining actin was differentially expressed in fertile and male-sterile floral samples, but not different across environments or interactions, all fertile samples were normalized to fertile actin relative quantity and male-sterile samples were normalized to male-sterile actin relative quantity. Relative expression for each technical replicate was calculated using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001), then the mean relative expression for each sample was calculated (Supplemental Table 3).

To determine if candidate genes were differentially expressed between fertile and male-sterile floral phenotypic classes or across environments, we analyzed the data using a split-plot experimental design in which whole-plots were temperature regimens across growth chamber and sub-plots were floral phenotypic classification within chambers (Wiebbecke et al. 2011a). PROC MIXED in SAS version 9.2 (SAS Institute Inc. 2008) was used to accommodate unbalanced fertile and male-sterile floral samples across temperature regimens. Temperature regimen (EnvID) and floral phenotypic classes (Pheno) were considered fixed effects and growth chamber was considered a random effect. PROC MIXED also was used to estimate the mean fold change and standard errors for each level combination of floral phenotypic classification and temperature regimen (EnvID×Pheno) for

each candidate gene (Supplemental Table 4). The mean fold change and standard errors also were calculated for each floral phenotypic classification within each temperature regimen (Supplemental Table 5). The interaction between floral phenotypic classification and temperature regimen were significant for Glyma03g31530 and Glyma03g30880. Therefore, we used contrast statements to test significance of temperature and photoperiod effects within floral phenotypic classes (Fig. 1 and 2).

RESULTS

Candidate Gene Identification

Using the soybean genome browser (www.soybase.org), we were able to visualize the *ms9* region between SSR markers Satt521 and Satt237 on chromosome 03 (Cervantes-Martinez et al. 2007). Once the genomic region was identified (www.soybase.org), all assigned gene annotations (Glyma03g30740 to Glyma03g03g32450) provided in SoyBase, and/or obtained by BLASTX ($E<10^{-20}$, Altschul et al. 1997) against the GenBank nonredundant database (nr), were used to search the available literature for possible involvement in nuclear encoded male sterility, or floral development, influenced by the environment. A combination of two different methods was used to narrow the candidate genes based on floral expression. First, candidate gene coding sequences were compared to the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005) using BLASTN ($E<10^{-6}$, Altschul et al. 1997). The soybean gene index is comprised of tentative consensus sequences (TCs) built from overlapping expressed sequence tags (ESTs). TCs identified by BLAST were examined for the presence of floral ESTs, as indicated by library identifiers (Gm-c1015, c1016, c1051, and c1061, Shoemaker et al. 2002). Second, gene

identifiers of potential candidates were queried against the recently published RNA-Seq atlas for soybean (Severin et al. 2010) to determine if candidate genes were significantly overrepresented with floral reads. We grouped raw (unnormalized) expression data by floral (open flowers) and nonfloral tissues (young leaf, one cm pod, pod shell 10 days after fertilization (DAF), pod shell 14 DAF, seed 10 DAF, seed 14 DAF, seed 21 DAF, seed 28 DAF, seed 35 DAF, seed 42 DAF, root, and nodule). A Fisher's exact test (Fisher 1949) was used to identify candidate genes significantly overrepresented with floral reads when compared to all other reads and relative to all genes in the genome with a Bonferoni correction for oversampling (P-value ≤ 0.05, Bonferroni 1936). After narrowing our candidate gene list to those with expression data supporting floral involvement, those genes were compared against The Arabidopsis Information Resource (TAIR, www.arabidopsis.org). We used TBLASTX ($E<10^{-6}$, Altschul et al. 2008) to compare soybean and Arabidopsis coding sequences and determine if Arabidopsis homologs were identified previously as male-sterility genes, were expressed in floral tissues, or had mutants with altered floral phenotypes.

Using this approach, we were able to identify nine candidate genes for further expression analyses (Table 1). Two of the genes encoded related small nuclear ribonucleoproteins (snRNPs). Glyma03g30880 (XP_002520764, 5.00 $E\times10^{-62}$) and Glyma03g30750 (NP_563682, 1.00 $E\times10^{-54}$) each have a LSm (like Sm) domain thought to be involved in spliceosome assembly (Will and Lührmann 2001). The spliceosome core structure is composed of small nuclear RNAs complexed with seven Sm proteins, all required for proper splicing of immature mRNAs. Environmental cues, such as abiotic stress and temperature, can effect splicing (Iida et al. 2004). In rice, defective splicing of UDP-glucose

pyrophosphorylase (*Ugp1*) results in male sterility. Since reversion to fertility occurs at low temperatures, Chen et al. (2007, 2009) proposed temperature sensitive splicing could be regulated by the spliceosomal complex. BLASTN (Altschul et al. 1997) comparison against the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005) revealed matches to TC427774 (Glyma03g30880, 4.60 $E\times10^{-73}$, 100% identity) and TC457314 (Glyma03g30750, 1.00 $E\times10^{-58}$, 100% identity), which were composed predominantly of ESTs from globular stage embryos. Queries of the RNA-Seq atlas (Severin et al. 2010) identified 37 and 95 raw floral reads for Glyma03g30880 and Glyma03g30750, respectively, however, neither gene was statistically overrepresented ($P<0.05$) with floral reads. The closest Arabidopsis homologs, AT4G02840 (TBLASTX, 1.00 $E\times10^{-57}$) and AT1G03330 (TBLASTX, 5.00 $E\times10^{-52}$) were also expressed in globular stage embryos, as well as male gametophytes, floral, and stamen tissues.

The third candidate gene, Glyma03g30770 had homology to isoamylase, (BAF52942, E-value = 0.00) a starch debranching enzyme that breaks starch into sugars used for plant growth and development (Stanley et al. 2005). After exposing Arabidopsis plants to cold treatment, amylase expression was not induced in the anther, however, amylase expression was highly induced in leaves (Lee and Lee 2003). These results were unexpected because amylases are involved in cold acclimation. Lee and Lee (2003) proposed poor accumulation of amylases resulted in pollen cold sensitivity. There were no TCs related (BLASTN, $E<10^{-6}$, Altschul et al. 1997) to Glyma03g30770 in the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005). Nine floral reads corresponding to Glyma03g30770 (not significantly floral) were identified from the RNA-Seq atlas (Severin et al. 2010). The closest Arabidopsis homolog, AT1G03310 (TBLASTX, E-value = 0.00) was expressed in

male gametophytes, floral, and stamen tissues. Mutations in AT1G03310 result in abnormal starch structures. *ms9* male-sterile pollen grains are devoid of starch (Wiebbecke et al. 2011a). Based on preliminary annotation, it is possible a mutation in Glyma03g30770, or its promoter, may cause increased breakdown of available starch.

A MYC basic helix-loop-helix (bHLH) transcription factor, Glyma03g30940 (ADO13282, 5.00 E×10^{-178}), was the fourth candidate gene identified. In Arabidopsis, mutants of the MYC-class transcription factor (*ABORTED MICROSPORE*, *ams*) are devoid of viable pollen (Sorensen et al. 2003). Another MYC transcription factor, *Antirrhinum Delila* (*DEL*), is required for flower-tube pigmentation (Schijlen et al. 2004). MYC transcription factors also regulate anthocyanin biosynthesis, as a part of the flavonoid biosynthesis pathway, and are involved in insect pollinator attraction (Schijlen et al. 2004). Given that Ortiz-Perez et al. (2006) demonstrated *ms9* lines had higher insect-mediated seed-set, a MYC transcription factor could be responsible for altering the regulation of anthocyanin resulting in greater insect pollinator attraction, as well as male-sterility. In the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005), Glyma03g30940 was most related to TC485385 (8.40 E×10^{-18}, 75% identity), which is composed exclusively of ESTs from globular stage embryos. Glyma03g30940 had only two raw floral reads in the RNA-Seq atlas (Severin et al. 2010) and was not significantly floral. Mutations in the closest Arabidopsis homolog, AT1G63650 (TBLASTX, 3.00 E×10^{-60}) resulted in reduced trichomes (hairs on the plant surface often referred to as pubescence), anthocyanin, and seed mucilage. AT1G63650 was expressed in male gametophytes and floral tissues.

The fifth and sixth candidate genes were a Miro GTPase and a GTP-binding protein, Glyma03g30990 (XP_002520752, E-value = 0.00) and Glyma03g31040 (XP_002520741, 4.00 E×10^{-116}), respectively. Miro GTPases are involved in the pollen-tube growth and germination (Krichevsky et al. 2007; Yamaoka and Leaver 2008). Krichevsky et al. (2007) showed there are a number of energy-dependent processes that effect pollen-tube germination and growth. In Arabidopsis, *MIRO1*, disrupts mitochondrial streaming, which may reduce energy, resulting in impaired pollen-tube germination and growth (Yamaoka and Leaver 2008). In the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005) greatest homology was found to TC455255 (Glyma03g30990, 2.50 E×10^{-105}, 96% identity) and TC453868 (Glyma03g31040, 1.20 E×10^{-152}, 95% identity), both composed exclusively of ESTs from globular stage embryos. Queries of the RNA-Seq atlas (Severin et al. 2010) identified 27 and 14 raw floral reads for Glyma03g30990 and Glyma03g31040, respectively, however, neither was significantly overrepresented with reads from the floral library. The closest Arabidopsis homologs, AT3G63150 (TBLASTX, E-value = 0.00) and AT4G02790 (TBLASTX, 1.00 E×10^{-109}) were expressed in floral, stamen, pollen, and male-gametophyte tissues.

The seventh candidate gene identified was a sugar transport protein, Glyma03g31020 (CAD61275, 4.00 E×10^{-29}). Sugar transporters are responsible for moving sugars from photosynthetically active (source) tissues to developing (sink) tissues, like pollen (Goetz et al. 2001). If sugar transporters were impaired in male-sterile plants, it could inhibit carbohydrate supplies, decreasing pollen germination efficiency. Glyma03g31020 shared greatest homology to TC441197 (1.10 E×10^{-13}, 63% identity) in the soybean gene index (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005). TC441197 was expressed in

root, hypocotyl, and floral tissues. Glyma03g31020 was not expressed in the RNA-Seq atlas (Severin et al. 2010). The closest Arabidopsis homolog, AT1G09960 (TBLASTX, 1.00 $E\times10^{18}$) was expressed in male gametophytes, stamen, and floral tissues.

The eighth candidate gene we identified encoded a calcium transporting ATPase, Glyma03g31420 (AAL17950, E-value = 0.00). This gene was selected because Ca^{2+} signals play an important role in pollen-tube growth and fertilization (Schiøtt et al. 2004). Cytosolic Ca^{2+} gradients are suggested to be involved in localized growth by altering Ca^{2+} signals (Frietsch et al. 2007). Frietsch et al. (2007) showed *cyclin nucleotide-gated channel 18* (*cngc18*) was localized to the pollen-tube tip and knock-out mutants have impaired pollen-tube growth. Glyma03g31420 had homology to TC424581 (3.30 $E\times10^{-245}$, 99% identity), which is composed primarily of ESTs from globular stage embryos (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005). Query of the RNA-Seq atlas (Severin et al. 2010) identified 141 raw floral reads corresponding to Glyma03g31420, making it significantly floral ($P < 0.05$). The closest Arabidopsis homolog, AT3G63380 (TBLASTX, E-value = 0.00) was expressed in male gametophytes, stamen, and floral plant tissues.

The last candidate gene identified for *ms9* was an auxin- indole-3-acetic acid (AUX/IAA) responsive gene, Glyma03g31530 (NP_001236693, 8.00 $E\times10^{-138}$). AUX/IAA responsive genes are transcription factors that play a role in negatively regulating auxin signaling (Mockaitis and Estelle 2008). Cheng and Zhao (2007) outlined the critical role auxin plays in floral development during floral organ patterning. Auxin also has been shown to affect pollen maturation, filament growth, and anther dehiscence (Cecchetti et al. 2008). Sakata et al. (2010) demonstrated auxin biosynthesis is negatively impacted by increased temperature in Arabidopsis and barley TGMS lines. Glyma03g31530 had matched

TC482930 (2.50 E×10^{-165}, 100% identity, Dana Faber Cancer Institute Gene Indices, Lee et al. 2005), predominantly composed of ESTs from globular stage embryos. Query of the RNA-Seq atlas (Severin et al. 2010) identified 287 raw floral reads corresponding to Glyma03g31530, making it significantly floral ($P < 0.05$). The closest Arabidopsis homolog, AT3G04730 (TBLASTX, 3.00 E×10^{-66}) was expressed in floral, stamen, and male gametophyte plant tissues.

Expression Analyses of Candidate Genes in Floral Samples

To determine if the candidate genes could contribute to the male-sterile *ms9* phenotype, it was necessary to evaluate candidate gene expression in multiple controlled environments. In total, 448 selfed progeny of T359H were grown in seven different environments. 56 plants were allowed to mature in each growth chamber with one of the following night/day temperature regimes: 12/30, 15/30, 18/30, 23/30, and 18/35 °C with day length 13 h at flowering, 21/30 and 17/35 °C with 14 h day length at flowering, and temperature regime 23/30 °C repeated once (Wiebbecke et al. 2011a). At approximately six weeks, which corresponded to the R2 stage, reproductive buds and flowers were evaluated for pollen phenotype visually and by microscopy to confirm fertile or male-sterile pollen phenotype (Wiebbecke et al. 2011a). Fertile and male-sterile, female-fertile plants were identified in all environments in the expected segregation ratio of 3 fertile plants: 1 male-sterile, female-fertile plant. However, the selfed seed-set of the male-sterile, female-fertile plants decreased as day temperature increased from 30 to 35 °C, suggesting *ms9* was a TGMS (Wiebbecke et al. 2011a). Therefore, we collected all available floral tissue (reproductive buds at anthesis and one day prior to anthesis) from each of the fertile and male-sterile, female-fertile plants grown in seven environments. Because of expected

segregation for the recessive *ms9* locus, there were more fertile samples than male-sterile, female-fertile samples. Therefore, a subset of floral samples was used for expression analysis including all 102 male-sterile, female-fertile and 118 fertile plants encompassing the range of fertile selfed seed-set in each of the seven environments (Supplemental Table 3).

RNA from floral samples was not pooled across fertile or male-sterile floral phenotypic classification nor were they pooled across environments. We chose to maintain individual sample identity to detect subtle differences in fertile or male-sterile floral expression across environments. In attempts to increase statistical power to detect differences in expression, RNA from 220 floral samples (102 male-sterile and 118 fertile) were set up in triplicate blocks to serve as technical replicates for qRT-PCR.

Prior to expression analyses, one of the most important considerations was primer and TaqMan® probe set design. Soybean has a duplicated genome (Schmutz et al. 2010), therefore, each candidate gene sequence was compared against the soybean genome using BLASTN ($E < 10^{-20}$, Altschul et al. 1997) to identify homeologous genes (Table 1). The homeolog sequences were aligned with the candidate genes to identify sequence differences and probe sets were designed to target the coding sequence (CDS) or 3' untranslated region (UTR) to ensure specificity (Supplemental Table 1). DNA Oligo templates were designed to correspond to the expected amplification product for each homeolog with potential to amplify by gene-specific probes. The homeolog templates were amplified with gene-specific probes during probe validation to determine probe specificity. Following this analysis, probe sets were considered gene-specific if no amplification of the homeolog template was detected after 40 cycle thresholds (Ct). All assays were confirmed highly specific except the Glyma03g30750 assay. Amplification of the Glyma03g30750 homeolog template was

detected at 35.6 Ct for the homeolog template at a molar ratio of one copy amplicon in 50 ng/µL corn genomic DNA. Corn genomic DNA served as a carrier molecule to compensate for any potential efficiency differences that may have resulted from using water alone (Pfaffl 2002). However, amplification of Glyma03g30750 itself ranged from 25.8 Ct for the 250 ng/µL to 34.1 Ct for the 0.97 ng/µL serial dilution samples, respectively. This suggests that almost all expression detected originates from Glyma03g30750, not its homeolog.

Once gene specific probe sets were designed, qRT-PCR assays were validated using standard curve validation procedures targeting a slope of -3.3 ± 0.3, which reflected an amplification efficiency of 90 to 110% (Applied Biosystems 2008). All candidate gene assays performed within ideal qRT-PCR standards except four assays (Glyma03g30940, 84.17% efficiency, slope -3.77, Glyma03g41420, 82.46% efficiency, slope -3.83, Glyma03g31530, 87.80% efficiency, slope -3.65, and Glyma03g30990, 87.17% efficiency, slope -3.67). Although the assay validation for these genes was slightly outside the target range, no further modifications of the probes could be made without sacrificing gene specificity.

To determine relative expression of candidate genes, we first calculated the relative quantity of actin (Glyma19g00850, the endogenous housekeeping reference gene) for each sample. Next used SAS version 9.2 (SAS Institute Inc. 2008) to analyze mean relative quantity of actin mRNA between floral phenotypic classification, environments, and the interaction of floral phenotypic classification by environment. There was no difference in relative quantity of actin across environments ($P = 0.5405$), interactions ($P = 0.1196$), or growth chambers (when temperature and photoperiod were held constant, $P = 0.0914$).

However, there was a difference in relative quantity of actin between floral phenotypic classes, (fertile, 29.91 and male-sterile, 15.82, $P < 0.0001$, Supplemental Table 4).

Although there were differences in efficiency and slope outside of the target range for qRT-PCR, we chose to analyze all data for the candidate genes using the same statistical method to estimate relative gene expression as a fold change on $\log_2$ scale. Because of the difference in mean relative quantity of actin mRNA, relative candidate gene expression of floral technical replicates was normalized by their respective male-sterile or fertile floral relative quantity of actin mRNA. For each candidate gene, the relative gene expression of technical replicates was calculated ($2^{-\Delta\Delta Ct}$ method, Livak and Schmittgen 2001) and then averaged for each sample. Using a split-plot experimental design in which whole-plots were temperature regimens across growth chamber and sub-plots were floral phenotypic classification within chambers (Wiebbecke et al. 2011a). We used SAS version 9.2 (SAS Institute Inc. 2008) to analyze candidate gene relative expression across floral phenotypic classifications, environments, and the interaction of floral phenotypic classification by environment (Supplemental Table 4). Only two candidate genes, Glyma03g31530 and Glyma03g30880 were differentially expressed when considering the floral phenotypic classification by environment interaction ($P = 0.0327$ and $P = 0.0697$, respectively, Supplemental Table 4). We used contrast statements to test significance of night and day temperature, and photoperiod effects within floral phenotypic classes. Neither night temperature nor photoperiod by floral phenotypic classification interaction effects were significant. However, changing day temperature from 30 to 35 °C significantly changed expression for both Glyma03g31530 and Glyma03g30880 in male-sterile floral phenotype ($P= 0.0148$ and $P=0.0257$, respectively) (Fig.1 and 2). In addition, Glyma03g31530

expression in the fertile floral phenotype also had significant day temperature effect (*P* =0.0408, Fig. 1). There was suggestive evidence that Glyma03g30880 expression in fertile floral phenotypes (*P*=0.0752, Fig. 2) also was influenced by day temperature. Therefore, Glyma03g30880 is the better candidate for *ms9*, because it expression was significantly increased in male-sterile floral samples but only suggestive in fertile floral samples as day temperature increased. In all cases the effect of increased day temperature was a slight increase in expression of the candidate gene on the $\log_2$ scale (12.42 to 13.19 Glyma03g31530 male-sterile floral expression, 12.52 to 13.11 Glyma03g31530 fertile floral expression, 13.03 to 13.56 Glyma03g30880 male-sterile floral expression, and 13.14 to 13.55 Glyma03g30880 fertile floral expression, Fig. 1 and 2).

DISCUSSION

In the molecular region identified by Cervantes-Martinez et al. (2007), we were able to identify candidate genes using a bioinformatic approach evaluating each gene for involvement in male sterility based on the available thermosensitive genic male-sterility literature. We narrowed the list of candidate genes by confirming their floral expression utilizing publicly available data from soybean ESTs (Dana Faber Cancer Institute Gene Indices, Lee et al. 2005) and the RNA-Seq atlas (Severin et al. 2010). Finally, we evaluated involvement in male sterility using TAIR (www.arabidopsis.org), which resulted in the prioritized list of nine candidate genes described above.

Expression analyses of floral tissue from the selfed progeny of T359H collected from controlled environments enabled us to determine if any of the candidate genes were differentially expressed in fertile versus male-sterile floral phenotypic classes or in response

to environment. The endogenous control gene, actin (Glyma19g00850) was differentially expressed across fertile and male-sterile floral phenotypes but was not influenced by environment or floral phenotypic classification by environment interaction (Supplemental Table 4). Therefore, when examining the expression of the candidate genes, we normalized the expression relative to fertile or male-sterile actin, as determined by pollen phenotype of the sample. Using raw data (unnormalized) had no effect on the results (data not shown). In these analyses, seven candidate genes were not differentially expressed between fertile and male-sterile floral phenotypes or influenced by the environment, and are no longer considered candidates for the *ms9* thermosensitive genic male-sterility locus. We were able to identify two candidate genes (snRNP, Glyma03g30880 and AUX/IAA responsive, Glyma03g31530) that were differentially expressed in both fertile and male-sterile floral samples as day temperature increased from 30 to 35 °C.

Glyma03g30880 (snRNP) had homology to a key component of the spliceosome core structure required for correct splicing of immature mRNAs (Will and Lührmann 2001). Chen et al. (2007, 2009) hypothesized the splicing machinery is temperature sensitive, which could lead to improper splicing resulting in male sterility. Our analyses demonstrate that expression of Glyma03g30880 increased with increasing day temperature. Increasing the amount of Glyma03g30880 protein could alter the balance of snRNPs making up the spliceosomal complex. In addition, increasing temperature could alter protein structure inhibiting its ability to interact with other splisosomal proteins. This could result in incorrectly spliced product(s) resulting in a male-sterile phenotype. We proposed that temperature sensitive splicing could be the mechanism leading to thermosensitive genic male-sterility for *ms9*.

Glyma03g31530 (an AUX/IAA responsive gene) is a negative regulator of auxin signaling. In the absence of auxin, AUX/IAA responsive proteins form heterodimers with auxin response factors (ARFs) and bind to the promoters of auxin regulated genes, repressing their expression. When auxin levels are high, AUX/IAA responsive proteins are targeted for degradation allowing ARFs to dimerize with each other and induce expression of auxin regulated genes (Mockaitis and Estelle 2008). As thermosensitive male-sterile Arabidopsis and barley lines were exposed to increased temperatures, auxin levels were specifically decreased in the anther (Sakata et al. 2010). If auxin levels were decreased in the anther, AUX/IAA responsive proteins would not be targeted for degradation leading to the repression of auxin-regulated genes. Since auxin signaling is required for the regulation of anther dehiscence and pollen maturation (Cecchetti et al. 2008), repression of auxin-regulated genes could lead to male sterility. Interestingly, Zeng et al. (2007) used a proteomics approach to evaluate differences of the cytoplasmic male-sterile line NJCMS2A and its maintainer in soybean. Four proteins were specific to anthers of the fertile maintainer line, including AUX28 (Glyma19g34380), the homeolog of Glyma03g31530. Glyma19g34380 and Glyma03g31530 share 90% nucleotide identity, further supporting a role for Glyma03g31530 in pollen development. Our analyses showed an increase in expression of Glyma03g31530 in male-sterile and fertile flowers, which would decrease auxin signaling in both male-sterile and fertile flowers. It is unclear how expression in both fertile and male-sterile flowers could result in two different pollen phenotypes.

The approach described here helped prioritize the list of candidate genes for *ms9* male sterility. In addition, the two remaining candidate genes (Glyma03g30880 and Glyma03g31530) can be further characterized to prove they are responsible for male sterility

at the *ms9* locus. Repeating the analyses specifically looking at pollen expression, instead of entire floral tissue, may provide more conclusive evidence that one of these candidate genes is responsible for male sterility. An alternative approach would be to re-evaluate the progeny of T359H with newly designed molecular markers near or within the candidate genes to determine if the marker segregates with the male-sterile floral phenotypic classification. Once the male-sterility gene is identified molecular markers could be used in a marker assisted selection (MAS) program to improve the efficiency of identifying male-sterile, female-fertile plants for use as the female parent in hybrid seed production.

Acknowledgements

This project was a joint contribution from Monsanto Co., the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 4403, and from the USDA Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, and supported by the Hatch Act and the State of Iowa. The mention of a trademark or proprietary product by Monsanto Co., Iowa State University, or the USDA, and the use of the name by Monsanto Co., Iowa State University, or the USDA does not imply its approval to the exclusion of other products that may be suitable. We would like to express our appreciation to the Custom Expression Team at Monsanto, more specifically Todd Ziegler for his assistance with TaqMan® probe design, floral RNA extraction, and qRT-PRC.

REFERENCES

Altschul, SF, TL Madden, AA Schaffer, J Zhang, Z Zhang, W Miller, and DJ Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Applied Biosystems (2008) Application note real-time PCR: Understanding Ct. USA

Applied Biosystems (2001) ABI PRISM 7700 sequence detection system user bulletin #2. USA

Bonferroni, CE (1936) Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 8:3-62

Cecchetti, V, MM Altamura, G Falasca, P Constantino, and M Cardarelli (2008) Auxin regulates anther dehiscence, pollen maturation, and filament elongation. Plant Cell 20:1760-1774

Cervantes-Martinez, I, M Xu, L Zhang, Z Huang, KK Kato, HT Horner, and RG Palmer (2007) Molecular mapping of male-sterility loci *ms2* and *ms9* in soybean. Crop Sci 47:374-379

Chen, R, Y Pan, Y Wang, L Zhu, and G He (2009) Temperature-sensitive splicing is an important molecular regulation mechanism of thermosensitive genic male sterility in rice. Chinese Sci Bull 54:2354-2362

Chen, R, X Zhao, Z Shao, Z Wei, Y Wang, L Zhu, J Zhao, M Sun, R He, and G He (2007) Rice *UDP-Glucose Pyrophosphorylase1* is essential for pollen callose deposition and its cosupression results in a new type of thermosensitive gene male sterility. Plant Cell 19:847-861

Cheng, Y and Y Zhao (2007) A role for auxin in flower development. J Integrative Plant Biol 49:99-104

Economic Research Service (2010) Soybean and oil crops briefing room: Overview. [www.ers.usda.gov/Briefing/SoybeansOilCrops/]. (verified on 02Oct2011). USDA, ERS

Fehr, WR and CE Caviness (1977) Stages of soybean development. Spec Rep 80. Iowa Agri Home Econ Expt Stn, Iowa State Univ, Ames, IA

Fisher, RA (1949) A preliminary linkage test with agouti and undulated mice; the fifth linkage-group. Heredity 3:229-241

Food and Agriculture Organization (2009) How to feed the world 2050: Global agriculture towards 2050. [www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_ Global_Agriculture.pdf]. (verified on 09Oct2011). FAO

Frietsch, S, YF Wang, C Sladek, LR Poulsen, SM Romanowsky, JI Schroeder, and JF Harper (2007) A *cyclin nucleotide-gated channel* is essential for polarized tip growth of pollen. PNAS 104:14531-14536

GenBank [blast.ncbi.nlm.nih.gov/Blast.cgi]

Goetz, M, DE Godt, A Guivarc'h, U Kahmann, D Chriqui, and T Roitsch (2001) Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. PNAS 98:6522-6527

Iida, KM, Seki, T Sakurai, M Satou, K Akiyama, T Toyoda, A Konagaya, and K Shinozaki (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. Nucleic Acids Res 32:5096-5103

Jenkins, JN, JC McCarty, J Wu, and OA Gutierrez (2009) Genetic variance components and genetic effects among eleven diverse upland cotton lines and their F2 hybrids. Euphytica 167:397-408

Jin, W, HT Horner, and RG Palmer (1997) Genetics and cytology of a new genic male-sterile soybean [*Glycine max* (L.) Merr.]. Sex Plant Reprod 10:13-21

Krichevsky, A, SV Kozlovsky, GW Tian, MH Chen, A Zaltsman, and V Citovsky (2007) How pollen tubes grow. Dev Biol 303:405-420

Lee, J, and D Lee (2003) Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. Plant Physiol 132:517-529

Lee, Y, J Tsai, S Sunkara, S Karamycheva, G Pertea, R Sultana, V Antonescu, A Chan, F Cheung, and J Quackenbush (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res 33:D71-D74

Li, J, and L Yuan (2000) Hybrid rice: Genetics, breeding, and seed production. Plant Breed Rev 17:15-158

Livak, KJ and TD Schmittgen (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402-408

McCarty, JC, JN Jenkins, and J Wu (2004a) Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. Crop Sci 44:1226-1230

McCarty, JC, JN Jenkins, and J Wu (2004b) Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: II. Genetic effects and genotypic values. Crop Sci 44:1231-1235

Mockaitis, K and M Estelle (2008) Auxin receptors and plant development: a new signaling paradigm. Annu Rev Cell Dev Biol 24:55-80

Ortiz-Perez, E, HT Horner, SJ Hanlin, and RG Palmer (2006) Insect-mediated seed-set evaluation of 21 soybean lines segregating for male sterility at 10 different loci. Euphytica 152:351-360

Palmer, RG, PT Perez, E Ortiz-Perez, F Maalouf, and MJ Suso (2009) The role of crop-pollinator relationships in breeding for pollinator-friendly legumes: from a breeding perspective. Euphytica 170:35-52

Palmer, RG, J Gai, H Sun, and JW Burton (2001) Production and evaluation of hybrid soybean. Plant Breed Rev 21:263-307

Palmer, RG (2000) Genetics of four male-sterile, female-fertile soybean mutants. Crop Sci 40:78-83

Pfaffl, MW (2002) Quantification strategies in real-time PCR. In: SA Bustin (ed) A-Z of quantitative PCR. Int Univ Line. La Jolla, CA, pp 87-112

Population Reference Bureau (2011) World population growth, 1950-2050. [www.prb.org/Educators/TeachersGuides/HumanPopulation/PopulationGrowth.aspx]. (verified on 02Oct2011). Washington, DC

Sakata, T, T Oshino, S Miura, M Tomabechi, Y Tsunaga, N Higashitani, Y Miyazawa, H Takahashi, M Watanabe, and A Higashitani (2010) Auxins reverse plant male sterility caused by high temperatures. PNAS 107:8569-8574

SAS Institute Inc (2008) SAS/STAT 9.2 user's guide. Cary, NC

Saxena, KB, R Sultana, N Mallikarjuna, RK Saxena, RV Kumar, SL Sawargaonkar, and RK Varshney (2010) Male-sterility systems in pigeonpea and their role in enhancing yield. Plant Breed 129:125-134

Schijlen, EGWM, CH Ric de Vos, AJ van Tunen, and AG Bovy (2004) Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648

Schiøtt, M, SM Romanowsky, L Bækgaard, MK Jakobsen, MG Palmgren, and JF Harper (2004) A plant plasma membrane Ca2+ pump is required for normal pollen tube growth and fertilization. PNAS 101:9502-9507

Schmutz, J, SB Cannon, J Schlueter, J Ma, T Mitros, W Nelson, DL Hyten, Q Song, JJ Thelen, J Cheng, D Xu, U Hellsten, GD May, Y Yu, T Sakurai, T Umezawa, MK Bhattacharyya, D Sandhu, B Valliyodan, E Lindquist, M Peto, D Grant, S Shu, D Goodstein, K Barry, M Futrell-Griggs, B Abernathy, J Du, Z Tian, L Zhu, N Gill, T Joshi, M Libault, A Sethuraman, X Zhang, K Shinozaki, HT Nguyen, RA Wing, P Cregan, J Specht, J Grimwood, D Rokhsar, G Stacey, RC Shoemaker, and SA Jackson (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183

Severin, AJ, JL Woody, Y Bolon, B Joseph, BW Diers, AD Farmer, GJ Muehlbauer, RT Nelson, D Grant, JE Spect, MA Graham, SB Cannon, GD May, CP Vance, and RC Shoemaker (2010) RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10:160-176

Shoemaker, R, P Keim, L Vodkin, E Retzel, SW Clifton, R Waterston, D Smoller, V Coryell, A Khanna, J Erpelding, X Gai, V Brendel, C Raph-Schmidt, EG Shoop, CJ Vielweber, M Schmatz, D Pape, Y Bowers, B Thelsing, J Martin, M Dante, and C Granger (2002) A compliation of soybean ESTs: generation and analysis. Genome 45: 329-338

SoyBase [www.soybase.org]

Sorensen, AM, S Kröber, US Unte, P Huijser, K Dekker, and H Saedler (2003) The Arabidopsis *ABORTED MICROSPORE* (*AMS*) gene encodes a MYC class transcription factor. Plant J 33:413-423

Specht, JE and JH Williams (1984) Contribution of genetic technology to soybean productivity – retrospect and prospect. In: WR Fehr (ed) Genetic contributions to yield gains of five major crop plants. Madison, WI, pp 49-74

Stanley, D, KJF Farnden, and EA MacRae (2005) Plant α-amylases: functions and roles in carbohydrate metabolism. Biologia 60: Suppl 16:65-71

TAIR [www.arabidopsis.org].

Virmani, SS, ZM Sun, TM Mou, AJ Ali, and CX Mao (2003) Two-line hybrid rice seed production. In: B Hardy (ed) Two-line hybrid rice breeding manual. Los Baños, Philippines, pp 53-64

Wiebbecke, CE, MA Graham, SC Cianzio, and RG Palmer (2011a) Day temperature influences the male-sterility locus *ms9* in soybean. Submitted to Crop Sci

Wiebbecke, CE, MA Graham, SC Cianzio, and RG Palmer (2011b) Computational identification of candidate genes in soybean male-sterility loci.

Will, CL and R Lührmann (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol 13:290-301

Yamaoka, S and CJ Leaver (2008) *EMB2473/MIRO1*, an Arabidopsis *Miro GTPase*, is required for embryogenesis and influences mitochondrial morphology in pollen. Plant Cell 20:589-601

Zeng, WY, SP Yang, DY Yu, and JY Gai (2007) A comparative study on anther proteomics between cytoplasmic-nuclear male-sterile line NJCMS2A and its maintainer of soybean. Acta Agronomica Sinica 33:1637-1643

Table 1. Male-sterility candidate genes for male-sterility locus *ms9* in soybean. BLAST (Altschul et al. 1997) results were obtained using predicted coding sequence for each candidate gene.

Functional Annotation	ms9 Glyma 1 assembly ID[a]	Homeolog to *ms9* Glyma 1 assembly ID[a]	GenBank BLASTX	GenBank E-value	Soybean gene index BLASTN (soybean)	Soybean gene index E-value	TAIR TBLASTX	TAIR E-value
snRNP core protein	Glyma03g30750	Glyma19g33600	NP_563682	1.00 E×10^{-54}	TC457314	1.00 E×10^{-58}	AT1G03330	5.00 E×10^{-52}
Isoamylase	Glyma03g30770	Glyma19g33620	BAF52942	0.00	TC425306	0.084	AT1G03310	0.00
snRNP	Glyma03g30880	Glyma19g33700	XP_002520764	5.00 E×10^{-62}	TC427774	4.60 E×10^{-73}	AT4G02840	1.00 E×10^{-57}
MYC	Glyma03g30940	Glyma19g33770	ADO13282	5.00 E×10^{-178}	TC485385	8.40 E×10^{-18}	AT1G63650	3.00 E×10^{-60}
Miro GTPase	Glyma03g30990	Glyma19g33810	XP_002520752	0.00	TC455255	2.50 E×10^{-105}	AT3G63150	0.00
Sucrose transport protein	Glyma03g31020	N/A[b]	CAD61275	4.00 E×10^{-29}	TC441197	1.10 E×10^{-13}	AT1G09960	1.00 E×10^{-18}
GTP binding protein	Glyma03g31040	Glyma19g33870	AP_002520741	4.00 E×10^{-116}	TC453868	1.20 E×10^{-152}	AT4G02790	1.00 E×10^{-109}
Calcium transporting ATPase	Glyma03g31420	Glyma19g34250	AAL17950	0.00	TC424581	3.30 E×10^{-245}	AT3G63380	0.00
AUX/IAA responsive	Glyma03g31530	Glyma19g34380	NP_001236693	8.00 E×10^{-138}	TC482930	2.50 E×10^{-165}	AT3G04730	3.00 E×10^{-66}

[a] Soybean sequence Glyma 1, Schmutz et al. (2010).

[b] No homeologous gene identified in homeologous region on chromosome Gm19.

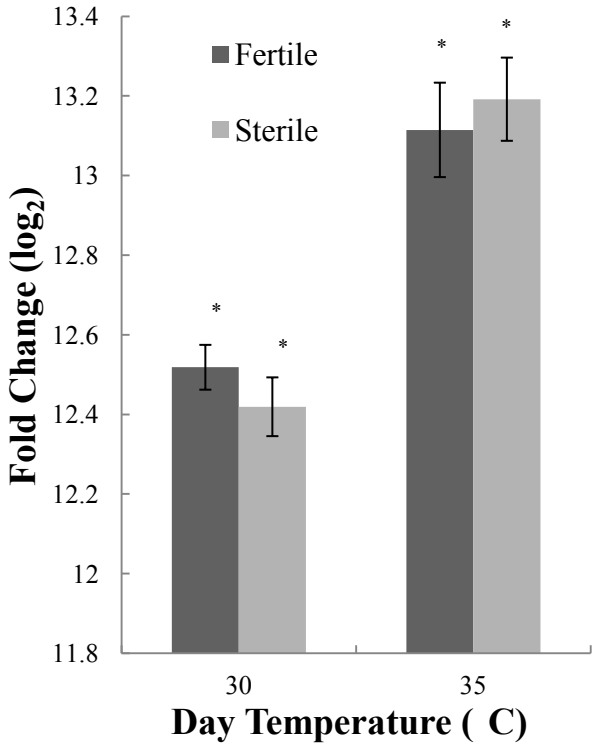

Figure 1. Glyma03g31530 (AUX/IAA responsive gene) mean relative expression expressed as a fold change (log_2) calculated by $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) in male-sterile (*ms9 ms9*) and fertile (*Ms9 Ms9* or *Ms9 ms9*) floral samples increased as day temperature increased from 30 to 35 °C, $P = 0.0148$ and $P = 0.0408$. Contrast statements were used to compare the effect of day temperature on floral expression of Glyma03g31530, error bars represent standard error of the mean relative expression calculated by PROC MIXED in SAS version 9.2 (SAS Institute Inc. 2008).

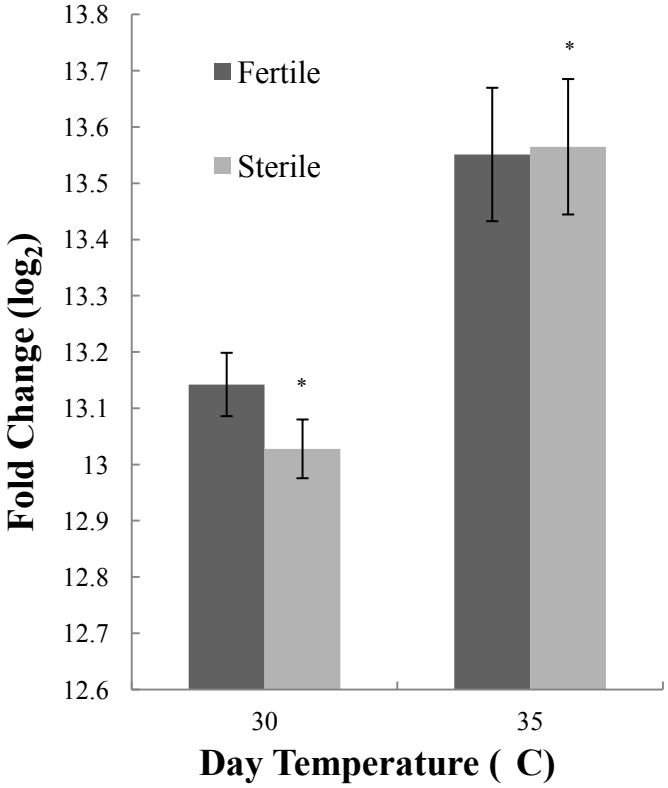

Figure 2. Glyma03g30880 (snRNP gene) mean relative expression expressed as a fold change ($\log_2$) calculated by $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) in male-sterile (*ms9ms9*) floral samples increased as day temperature increased from 30 to 35 °C, P = 0.0257. There was suggestive although inconclusive evidence that the increase in snRNP gene expression in fertile (*Ms9Ms9* or *Ms9ms9*) floral samples was also influenced by day temperature (P = 0.0752). Error bars represent standard error of the mean relative expression calculated using contrast statements to test day temperature effect on Glyma03g30880 floral expression in PROC MIXED (SAS version 9.2, SAS Institute Inc. 2008).

CHAPTER 5: CONCLUSIONS

Our results from Chapter 2 will facilitate the development of protocols to utilize the *ms9* source of male sterility in soybean breeding and hybrid seed production programs, similar to those outlined in the Two-line hybrid rice breeding manual (Virmani et al., 2003). In our work, we concluded that *ms9* exposed to high day temperature (35 °C) favored male sterility, whereas lower day temperature, about 30 °C, resulted in partial reversion to male fertility. These results contribute to address the fifth requirement outlined by Palmer et al. (2001) for hybrid seed production, which was 'to economically increase seed for the successful implementation of hybrid seed production in soybean'. Production seed growers can utilize controlled temperature environments (field hoop houses, greenhouses, or growth chambers), while eliminating insect pollinators, to develop systems conducive to male-sterile line maintenance. Additionally, soybean breeders can manipulate temperatures to limit self-pollination, and encourage natural outcrossing by supplementing environments with insect pollinators to generate hybrid seed or to facilitate more efficient manual cross-pollination. Determining an efficient method and a defined protocol with appropriate environmental coordinates for maximum and consistent expression of the *ms9* male-sterility gene will facilitate production of hybrid seed on a broad scale, enabling yield tests in multiple environments. Additionally, diverse germplasm combinations may be used in hybrid seed production in attempts to identify heterotic groups and yield heterosis expression in soybean. Establishment of the critical time of day temperature may lead to optimization of hybrid seed production systems.

In Chapter 3 we applied a bioinformatic approach to identify candidate genes related to male sterility. Our analysis identified 23 candidate genes in three of the four mapped

environmentally sensitive male-sterility loci (*ms9*, *msp*, and *ms3*) and the environmentally stable *ms2* locus. The computational analysis also identified transcription factor binding sites (TFBS) common across multiple environmentally sensitive male-sterility loci, as well as TFBS that were unique to each male-sterility locus. This initial identification of candidate male-sterility genes in soybean, accomplished for the first time, will allow future experiments to elucidate the functional aspects of candidate male-sterility genes. We also have demonstrated that bioinformatic approaches used enables researchers identify candidate genes for any mapped trait of interest. The genomic tools available for model crops and soybean provide novel resources to analyze genes and determine putative annotations. Further conformational testing can be pursued for a detailed subset of candidate genes for any trait of interest.

Similarly, in Chapter 4, with the identification of specific genes with differential expression among male-sterile, female-fertile plants as day temperature increases also will permit further analysis to confirm male-sterility expression in relation to environmental factors. A traditional molecular approach can be used, in which new molecular markers, simple sequence repeats (SSRs) or single nucleotide polymorphisms (SNPs), can be designed for each candidate gene within a male-sterility locus. The markers can be screened against segregating populations to determine if they cosegregate with plants displaying the male-sterile phenotype. A different modern technique to determine function and phenotypic association of candidate gene(s) would be to utilize RNA interference (RNAi, Libault et al., 2009) or virus induced gene silencing (VIGS, Zhang et al., 2009) to silence expression of the candidate gene(s). The silenced plants would be characterized in attempts to identify phenotypes associated with male sterility. Once candidate genes are confirmed, polymorphic

markers can be designed from within or near the gene, and marker assisted selection (MAS) programs could be used to identify male-sterile, female-fertile plants prior to flowering. Ultimately, this approach would greatly improve the efficiency of identifying plants for use as female parents in hybrid seed production.

APPENDIX A

Chapter 3. Supplemental Table 1. Plant transcription factor binding sites (TFBS) identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006) for *ms9* locus versus *ms9*H locus. Total number of genes in each dataset is in parenthesis below each dataset name in the heading.

Plant TFBS Name	*ms9*A (150)[a]	*ms9*F (123)[b]	*ms9*SF (35)[c]	*ms9*HA (153)[a]	*ms9*HF (143)[b]	*ms9*HSF (40)[c]
ABF_Q2	52	43	10	50 *	41 *	17 *
ABF1_02	29 *	25 *	4	21	19	6
ABI4_01	54	42	12	82	77 *	26
ABZ1_01	130 *	112	33	143	128	48
AG_01	104 *	91 *	28	75	67	18
AG_02	81	75	23	88 *	77 *	22
AG_03	104 *	91 *	28	75	67	18
AGL1_01	124 *	100 *	24	92	83	21
AGL1_02	116 *	94	22	82	74	18
AGL15_01	140 *	126 *	32	115	95	20
ALFIN1_Q2	376	324	82	349 *	309 *	98
ATMYB77_01	52	46 *	14	56 *	51 *	14
BPC1_Q2	133	108	27	144	131 *	39
BZIP911_02	39 *	27	7	36	29	12 *
CG1_Q6	200 *	168 *	52	195 *	176 *	48
CPRF_Q2	138 *	119 *	40 *	133 *	116 *	40 *
CPRF1_01	82 *	70 *	19 *	90 *	77 *	23 *
CPRF2_01	79 *	64 *	24	84	72	22
CPRF2_Q2	78 *	62 *	22	87 *	74 *	25 *
CPRF3_01	81 *	64	24 *	85 *	75 *	23 *
CPRF3_Q2	87 *	71 *	25 *	84 *	73 *	22 *
DOF_Q2	263	215 *	58	279 *	253 *	87 *
E2F_Q2	104 *	94 *	28	105 *	89 *	19
EMBP1_02	123	103	29 *	120	105	36 *
EMBP1_Q2	105 *	92 *	34	100 *	82 *	35 *
GBF_Q2	113	91	37	116	105	36 *
GBP_Q6	154 *	124	48	170	150	58 *
HAHB4_01	95	81	26 *	121 *	107 *	29
HBP1A_Q2	139 *	118 *	45	150 *	130 *	44 *
KNOX3_01	175	146 *	46	172 *	151 *	39
MADSA_Q2	160	140 *	36	166	146	34
O2_02	117 *	93 *	35	127	109	40
OCSBF1_01	169	139	45*	191 *	169 *	52
PBF_01	195	163 *	52	199	181	60
PIF3_01	53	44 *	9	54	47	18

APPENDIX A

Chapter 3. Supplemental Table 1 (continued).

Plant TFBS Name	***ms9*A (150)[a]**	***ms9*F (123)[b]**	***ms9*SF (35)[c]**	***ms9*HA (153)[a]**	***ms9*HF (143)[b]**	***ms9*HSF (40)[c]**
PIF3_02	51 *	46 *	6	47	41	19 *
RAV1_02	123 *	94	24	141	126	31
RITA1_01	127	102	39 *	145	124	39
SED_Q2	180	151	37	179	163	65 *
TAF1_01	196 *	162 *	54	218 *	188 *	54
TAF1_Q2	154 *	125 *	49	158 *	137 *	43 *
TGA1B_01	91 *	69	24	91 *	79 *	25
TGA1B_Q2	154 *	123	39	167 *	146	46
TRAB1_Q2	184 *	148 *	32	178	159	40

[*] Significantly overrepresented TFBS, $t \leq 0.05$.

[a] All genes dataset.

[b] Genes with at least 1 floral read in RNA-Seq atlas (Severin et al., 2010).

[c] Genes with floral expression significantly overrepresented determined using Fisher's exact test (Fisher, 1949) with Bonferroni correction (P-value ≤ 0.05, Bonferroni, 1936).

APPENDIX A

Chapter 3. Supplemental Table 2. Plant transcription factor binding sites (TFBS) identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006) for *ms8* locus versus *ms8*H locus. Total number of genes in each dataset is in parenthesis below each dataset name in the heading.

Plant TFBS Name	***ms8*A (14)[a]**	***ms8*F (11)[b]**	***ms8*SF (6)[c]**	***ms8*HA (20)[a]**	***ms8*HF (13)[b]**	***ms8*HSF (7)[c]**
AGP1_01	13 *	10	7 *	17	12	4
ARR10_01	10	7	3	21	14 *	5
BHLH66_01	16	11	9 *	20	13	9 *
C1_Q2	15	12	5	30 *	20 *	11 *
DOF1_01	22	19 *	6	26	15	9
DOF2_01	18	13	7	29 *	18	8
DOF3_01	18	12	4	32 *	22 *	13 *
GAMYB_01	16	10	5	21	16 *	7 *
HAHB4_01	5	0	0	12	8	6 *
LIM1_01	25	13	4	50 *	36 *	23 *
MYBAS1_01	20	17	5	28	18 *	10 *
RAV1_02	12	8	5	25 *	11	3
TEIL_01	7	5	2	18	13	8 *

[*] Significantly overrepresented TFBS, $t \leq 0.05$.

[a] All genes dataset.

[b] Genes with at least 1 floral read in RNA-Seq atlas (Severin et al., 2010).

[c] Genes with floral expression significantly overrepresented determined using Fisher's exact test (Fisher, 1949) with Bonferroni correction (P-value ≤ 0.05, Bonferroni, 1936).

APPENDIX A

Chapter 3. Supplemental Table 3. Plant transcription factor binding sites (TFBS) identified using Clover analysis (Frith et al., 2004) from TRANSFAC (Matys et al., 2006) for *ms2* locus versus *ms2*H locus. Total number of genes in each dataset is in parenthesis below each dataset name in the heading.

Plant TFBS Name	***ms2*A (373)[a]**	***ms2*F (341)[b]**	***ms2*SF (98)[c]**	***ms2*HA (510)[a]**	***ms2*HF (403)[b]**	**ms2HSF (110)[c]**
ABF_Q2	108 *	103 *	31 *	157	132	20
ABF1_02	59 *	58 *	22 *	74	63	9
ABI4_01	181 *	171 *	54	243 *	218 *	54 *
ABZ1_01	329 *	309 *	96 *	428 *	372 *	85
AG_01	202 *	185 *	59 *	256	215	57
AG_03	202 *	185 *	59 *	256	215	57
AGL1_01	250 *	236 *	74 *	350	290 *	75 *
AGL1_02	253 *	241 *	67 *	372	315	82
AGL2_01	222	202	58 *	261	209	63
AGP1_01	183 *	171 *	50 *	245	198	46
ALFIN1_Q2	835	793 *	229	1157 *	983 *	217
ATHB5_01	348	325	83 *	432	337	99
ATMYB15_01	41 *	38 *	17 *	61 *	49 *	6
ATMYB84_01	64 *	61 *	18 *	69	63	12
BHLH66_01	378	363	111	467	386 *	85
BZIP910_02	102	94	26 *	117	94	24
BZIP911_01	54	51	19 *	86	71	13
BZIP911_02	93	90	33 *	141	117	31
C1_Q2	462 *	432 *	135 *	580 *	478 *	147 *
CG1_Q6	488 *	466 *	154 *	651 *	551 *	136
CPRF_Q2	335	312 *	98 *	494 *	409 *	98
CPRF1_01	207 *	193 *	62 *	304 *	254 *	61
CPRF2_01	202 *	194 *	66 *	270 *	219 *	53
CPRF2_Q2	212 *	201 *	71 *	282 *	233 *	56
CPRF3_01	203 *	194 *	78 *	283 *	235 *	56
CPRF3_Q2	211	201 *	78 *	280	232 *	55
DOF3_01	467	439 *	143 *	662	556	141
E2F_Q2	241	224	51	309	243 *	64
EMBP1_02	303	279	78	388 *	314 *	72
EMBP1_Q2	259	242 *	83 *	364 *	298 *	66
ERF2_01	289 *	272 *	89 *	335	276	67
GAMYB_01	363 *	318	102	505	410	122 *
GBF_Q2	293 *	280 *	92 *	410 *	344 *	76
GBP_Q6	414 *	400 *	160 *	534 *	450 *	111
GT1_Q6	170 *	155 *	53 *	187	147	37

APPENDIX A

Chapter 3. Supplemental Table 3. (continued)

Plant TFBS Name	***ms2*A (373)[a]**	***ms2*F (341)[b]**	***ms2*SF (98)[c]**	***ms2*HA (510)[a]**	***ms2*HF (403)[b]**	**ms2HSF (110)[c]**
HBP1A_Q2	316	300	117 *	457	389 *	94
HBP1B_Q6	307	286 *	92 *	389 *	332 *	74
HBPA1_Q6_01	345	324 *	103 *	491 *	428 *	87
KNOX3_01	390 *	359 *	107 *	546 *	447 *	107
LIM1_01	718 *	672 *	205	938 *	789 *	219 *
MYBAS1_01	399 *	356 *	95	522 *	438 *	117
MYBPH3_02	119	104	32	190	147 *	45
O2_02	280 *	265 *	98 *	438	358 *	80
O2_Q2	292 *	279 *	76 *	364	302	81
OCSBF1_01	434 *	398 *	126 *	593 *	490 *	130 *
OSBZ8_Q6	386 *	371 *	115 *	445	382	94
P_01	333	311	100	537 *	434 *	128
PCF2_01	354 *	325 *	94	499	427 *	107
PCF5_01	271 *	255 *	75 *	315	276 *	76
PEND_01	238	218	58 *	397	304	77
PIF3_01	134	131 *	42 *	199	166	37
PIF3_02	131 *	128 *	32 *	181 *	150 *	24
RAV1_01	533 *	492 *	138	694 *	533 *	142
RAV1_02	351 *	321 *	95 *	386	314	74
RITA1_01	363 *	341 *	117 *	499 *	398 *	100
ROM_Q2	455 *	422 *	120 *	607 *	516 *	145
TAF1_01	505 *	477 *	167 *	648 *	540 *	123
TAF1_Q2	383	363 *	132 *	507	420 *	88
TEIL_01	263 *	232	55	349	275	84
TGA1A_Q2	388 *	371 *	136 *	536 *	443 *	99
TGA1A_Q2_01	256	239	73 *	368 *	302 *	79
TGA1B_01	217 *	209 *	74 *	278	232 *	54
TGA1B_Q2	409 *	391 *	150 *	547 *	462 *	100
TRAB1_Q2	458 *	441 *	143 *	563 *	488 *	105
WRKY_Q2	294 *	266 *	71	391	311	84
ZAP1_01	162	156 *	44 *	225	172	51

* Significantly overrepresented TFBS, $t \leq 0.05$.

[a] All genes dataset.

[b] Genes with at least 1 floral read in RNA-Seq atlas (Severin et al., 2010).

[c] Genes with floral expression significantly overrepresented determined using Fisher's exact test (Fisher, 1949) with Bonferroni correction (P-value ≤ 0.05, Bonferroni, 1936).

APPENDIX A

Chapter 3. Supplemental Text. *ms3* partial homolog gene range from Glyma1 genome assembly (Schmutz et al., 2010).

Glyma01g05320	Glyma01g05700	Glyma01g06080	Glyma01g06520
Glyma01g05330	Glyma01g05710	Glyma01g06090	Glyma01g06530
Glyma01g05340	Glyma01g05720	Glyma01g06100	Glyma01g06540
Glyma01g05350	Glyma01g05730	Glyma01g06110	Glyma01g06550
Glyma01g05360	Glyma01g05740	Glyma01g06120	Glyma01g06560
Glyma01g05370	Glyma01g05750	Glyma01g06130	Glyma01g06570
Glyma01g05380	Glyma01g05760	Glyma01g06140	Glyma01g06580
Glyma01g05390	Glyma01g05770	Glyma01g06150	Glyma01g06590
Glyma01g05400	Glyma01g05780	Glyma01g06160	Glyma01g06600
Glyma01g05410	Glyma01g05790	Glyma01g06170	Glyma01g06610
Glyma01g05420	Glyma01g05800	Glyma01g06180	Glyma01g06620
Glyma01g05430	Glyma01g05810	Glyma01g06250	Glyma01g06630
Glyma01g05440	Glyma01g05820	Glyma01g06260	Glyma01g06640
Glyma01g05450	Glyma01g05830	Glyma01g06270	Glyma01g06650
Glyma01g05460	Glyma01g05840	Glyma01g06280	Glyma01g06660
Glyma01g05470	Glyma01g05850	Glyma01g06290	Glyma01g06670
Glyma01g05480	Glyma01g05860	Glyma01g06300	Glyma01g06680
Glyma01g05490	Glyma01g05870	Glyma01g06310	Glyma01g06690
Glyma01g05500	Glyma01g05880	Glyma01g06320	Glyma01g06700
Glyma01g05510	Glyma01g05890	Glyma01g06330	Glyma01g06710
Glyma01g05520	Glyma01g05900	Glyma01g06340	Glyma01g06720
Glyma01g05530	Glyma01g05910	Glyma01g06350	Glyma01g06730
Glyma01g05540	Glyma01g05920	Glyma01g06360	Glyma01g06740
Glyma01g05550	Glyma01g05930	Glyma01g06370	Glyma01g06750
Glyma01g05560	Glyma01g05940	Glyma01g06380	Glyma01g06760
Glyma01g05570	Glyma01g05950	Glyma01g06390	Glyma01g06770
Glyma01g05580	Glyma01g05960	Glyma01g06400	Glyma01g06780
Glyma01g05590	Glyma01g05970	Glyma01g06410	Glyma01g06790
Glyma01g05600	Glyma01g05980	Glyma01g06420	Glyma01g06800
Glyma01g05610	Glyma01g05990	Glyma01g06430	Glyma01g06810
Glyma01g05620	Glyma01g06000	Glyma01g06440	Glyma01g06820
Glyma01g05630	Glyma01g06010	Glyma01g06450	Glyma01g06830
Glyma01g05640	Glyma01g06020	Glyma01g06460	Glyma01g06840
Glyma01g05650	Glyma01g06030	Glyma01g06470	Glyma01g06850
Glyma01g05660	Glyma01g06040	Glyma01g06480	Glyma01g06860
Glyma01g05670	Glyma01g06050	Glyma01g06490	Glyma01g06870
Glyma01g05680	Glyma01g06060	Glyma01g06500	Glyma01g06880
Glyma01g05690	Glyma01g06070	Glyma01g06510	Glyma01g06890

APPENDIX A

Chapter 3. Supplemental Text (continued)

Glyma01g06900	Glyma01g07380	Glyma01g07770	Glyma01g08220
Glyma01g06910	Glyma01g07390	Glyma01g07780	Glyma01g08230
Glyma01g06920	Glyma01g07400	Glyma01g07790	Glyma01g08240
Glyma01g06930	Glyma01g07410	Glyma01g07800	Glyma01g08250
Glyma01g06940	Glyma01g07420	Glyma01g07810	Glyma01g08260
Glyma01g06950	Glyma01g07430	Glyma01g07820	Glyma01g08540
Glyma01g06960	Glyma01g07440	Glyma01g07830	Glyma01g08550
Glyma01g06970	Glyma01g07450	Glyma01g07840	Glyma01g08560
Glyma01g06980	Glyma01g07460	Glyma01g07850	Glyma01g08570
Glyma01g06990	Glyma01g07470	Glyma01g07860	Glyma01g08580
Glyma01g07000	Glyma01g07480	Glyma01g07870	Glyma01g08590
Glyma01g07010	Glyma01g07490	Glyma01g07880	Glyma01g08600
Glyma01g07020	Glyma01g07500	Glyma01g07890	Glyma01g08610
Glyma01g07030	Glyma01g07510	Glyma01g07900	Glyma01g08980
Glyma01g07040	Glyma01g07520	Glyma01g07910	Glyma01g08990
Glyma01g07050	Glyma01g07530	Glyma01g07920	Glyma01g09000
Glyma01g07060	Glyma01g07540	Glyma01g07930	Glyma01g09010
Glyma01g07070	Glyma01g07550	Glyma01g07940	Glyma01g09020
Glyma01g07080	Glyma01g07560	Glyma01g07950	Glyma01g09030
Glyma01g07090	Glyma01g07570	Glyma01g07960	Glyma01g09040
Glyma01g07100	Glyma01g07580	Glyma01g07970	Glyma01g09050
Glyma01g07110	Glyma01g07590	Glyma01g07980	Glyma01g09060
Glyma01g07120	Glyma01g07600	Glyma01g07990	Glyma01g09070
Glyma01g07130	Glyma01g07610	Glyma01g08000	Glyma01g09080
Glyma01g07140	Glyma01g07620	Glyma01g08010	Glyma01g09090
Glyma01g07150	Glyma01g07630	Glyma01g08020	Glyma01g09100
Glyma01g07160	Glyma01g07640	Glyma01g08030	Glyma01g09110
Glyma01g07170	Glyma01g07650	Glyma01g08040	Glyma01g09120
Glyma01g07180	Glyma01g07660	Glyma01g08050	Glyma01g09130
Glyma01g07190	Glyma01g07670	Glyma01g08060	Glyma01g09140
Glyma01g07200	Glyma01g07680	Glyma01g08070	Glyma01g09150
Glyma01g07210	Glyma01g07690	Glyma01g08080	Glyma01g09160
Glyma01g07220	Glyma01g07700	Glyma01g08150	Glyma01g09170
Glyma01g07230	Glyma01g07710	Glyma01g08160	Glyma01g09180
Glyma01g07330	Glyma01g07720	Glyma01g08170	Glyma01g09190
Glyma01g07340	Glyma01g07730	Glyma01g08180	Glyma01g09200
Glyma01g07350	Glyma01g07740	Glyma01g08190	Glyma01g09210
Glyma01g07360	Glyma01g07750	Glyma01g08200	Glyma01g09220
Glyma01g07370	Glyma01g07760	Glyma01g08210	Glyma01g09230

APPENDIX A

Chapter 3. Supplemental Text (continued)

Glyma01g09240	Glyma01g10140	Glyma01g22600
Glyma01g09250	Glyma01g10150	Glyma01g22610
Glyma01g09260	Glyma01g22230	Glyma01g22620
Glyma01g09270	Glyma01g22240	Glyma01g22630
Glyma01g09280	Glyma01g22250	Glyma01g22640
Glyma01g09290	Glyma01g22260	Glyma01g22650
Glyma01g09300	Glyma01g22270	Glyma01g22660
Glyma01g09310	Glyma01g22280	Glyma01g22670
Glyma01g09320	Glyma01g22290	Glyma01g22990
Glyma01g09330	Glyma01g22300	Glyma01g23000
Glyma01g09340	Glyma01g22310	Glyma01g23010
Glyma01g09350	Glyma01g22320	Glyma01g23020
Glyma01g09360	Glyma01g22330	Glyma01g23030
Glyma01g09370	Glyma01g22340	Glyma01g23040
Glyma01g09380	Glyma01g22350	Glyma01g23050
Glyma01g09390	Glyma01g22360	Glyma01g23060
Glyma01g09400	Glyma01g22370	Glyma01g23070
Glyma01g09410	Glyma01g22380	Glyma01g23080
Glyma01g09420	Glyma01g22390	Glyma01g23090
Glyma01g09430	Glyma01g22400	Glyma01g23100
Glyma01g09440	Glyma01g22410	Glyma01g23110
Glyma01g09450	Glyma01g22420	Glyma01g23120
Glyma01g09460	Glyma01g22430	Glyma01g23130
Glyma01g09470	Glyma01g22440	Glyma01g23140
Glyma01g09480	Glyma01g22450	Glyma01g23150
Glyma01g09490	Glyma01g22460	Glyma01g23160
Glyma01g09500	Glyma01g22470	Glyma01g23170
Glyma01g09510	Glyma01g22480	Glyma01g23180
Glyma01g09520	Glyma01g22490	Glyma01g23190
Glyma01g10040	Glyma01g22500	Glyma01g23200
Glyma01g10050	Glyma01g22510	Glyma01g23210
Glyma01g10060	Glyma01g22520	Glyma01g23220
Glyma01g10070	Glyma01g22530	Glyma01g23230
Glyma01g10080	Glyma01g22540	
Glyma01g10090	Glyma01g22550	
Glyma01g10100	Glyma01g22560	
Glyma01g10110	Glyma01g22570	
Glyma01g10120	Glyma01g22580	
Glyma01g10130	Glyma01g22590	

Chapter 4. Supplemental Table 1. *ms9* male-sterility candidate gene assay primers and TaqMan® probes.

Assay Name	Glyma 1 ID[a]	Sequence type	Forward primer sequence	Reverse primer sequence	Probe sequence	Amplicon sequence	Amplicon Length (bp)	Tm	Homeolog template sequence
Gm ActinISU	Glyma19g00850	CDS	CGAGACCAGCTCTTCAGTGGAGAA	ACGCTCAGCGCCAATAGTGATAA	ATGAGTTGCCTGATGGGC	CGAGACCAGCTCTTCAGTGGAGAAGAGCTATGAGTTGCCTGATGGGCAGGTTATCACTATTGGCGCTGAGCGT	73	81	N/A
Amylase_2	Glyma03g30770	CDS	GGAAGTTGTTTTCTCCAATACCGCT	CACCAACTCCATTTGCATAATAGTAGGTT	AGATTGGTGCTCTACAGGGA	GGAAGTTGTTTTCTCCAATACCGCTGAGATTGGTGCTCTACAGGGAATTGACGACTCAACCTACTATTATGCAAATGGAGTTGGTG	86	78	GGAAGTTGTTTTCTCCAATACCGCAGAGATTGGTGCTATACAGGGAATTGATGACTCGTCCTACTATTATGCAAATGGAGTTGGTG
AUX_IAA	Glyma03g31530	CDS	CACTGAAGAAGTGCTCATCAGGAAGA	GATCTGCAGCAGCAGCAGCTT	ACTGAGGAGGATGAGTCTGCTA	CACTGAAGAAGTGCTCATCAGGAAGAGGAGTTTCTCTGAGACTGAAACTGAAACTGAGGAGGATGAGTCTGCTACCACTGTGGATTTGATGCTTAATCTTTCTTCTAAGGAAGCTGCTGCTGCAGATC	131	78	CACTGAAGAAGTGCTCATCAGGAAGAGGGGTTTCTCTGAGACTGAAACTGGTCATGAAGATGAGTCTGCCACCACTGTGGATTTGATGCTTAATCTTTCTTCTAAGGAGGCCGCAACCACTGC
Ca^{2+} Transport_3	Glyma03g31420	CDS	GTGGGGGATGTTGTGTCACTTAAG	TGATTCATCCACTTGCAAAGAATGA	ATTCCAGCTGATGGATTGTT	GTGGGGGATGTTGTGTCACTTAAGATTGGTGATCAAATTCCAGCTGATGGATTGTTCTTAAGTGGTCATTCTTTGCAAGTGGATGAATCA	90	76	GTGGGAGATATTGTATCACTTAAAATTGGTGATCAAATTCCAGCTGATGGGTTGTTCTTAAGTGGCTATTCTTTGCTAGTGGATGAATCA
GTP Binding Protein_3	Glyma03g31040	CDS	ATGTGATGACATTGGAGAGAGGTCCTAT	CTTTGCCTATTGTGGGAAGCTTTGT	ATGTGGCTGATGTTGCTGCA	ATGTGATGACATTGGAGAGAGGTCCTATGATGTGGCTGATGTTGCTGCAATTCTTGTGCAGATGCTTACAAAGCTTCCCACAATAGGCAAAG	92	78	ATGTGATGACATTGGAGAGAGGTCCTACGATGTGGCCGATGTTGCTGCAATTCTTGTGCAGATGCTTTCGAAGCTTCCCACAGTAGGTATGT
MYC	Glyma03g30940	CDS	CATACCCTCTTGTCACTGCCAATAATACTA	GTTAGGGACTGCTAATTCTGCTTCTTTG	CATTGTTGATAGAGGAGAAGCT	CATACCCTCTTGTCACTGCCAATAATACTATTCCAATCCAAAACATTGTTGATAGAGGAGAAGCTATAATATTGAATAATAATACCAAAGAAGCAGAATTAGCAGTCCCTAAC	113	74	N/A
Sucrose_2	Glyma03g31020	CDS	TAGCCGCTTCGACCGTCGTA	TGTCACTGAGGAGCCAACTGATG	CCTTATCATTGCTTACACCGCC	TAGCCGCTTCGACCGTCGTAGGCCCTTCATCCTCATCGATGTCGTCATCATCTTTGTCGTTGTCCTTATCATTGCTTACACCGCCAACATCAGTTGGCTCCTCAGTGACA	110	80	N/A
SnRNP_3	Glyma03g30880	Transcribed	GACGTTGCAGGAGTGCTGTTGTTAGTTAG	TTTCTTTCAGTGGAGACTTGAGAGATATGG	CTTGCATCAATAATTTGAAGTTCTCG	GACGTTGCAGGAGTGCTGTTGTTAGTTAGTCATGCAGACTTTGTTCATGTTGCCATTTTCTTGCATCAATGCTTTGAAGTTCTCGTGCATCAATAATTTGAAGTTCTCGTGTATAATTCATTCCTTCCTGTTGGATAATTGCCATATCTCTCAAGTCTCCACTGAAAGAAA	143	75	GACGTTGTAGGAGTGCTGTTGTTAGTTAATCATGCAGACTTTGTTCATGTTGCCATTTTCTTGCATCAATGCTTTGAAGTTCTCGTGTATAATTC-CTTCCTGTTGGATAATTGTCATATCTCTAAAGTCTCTTCTGAAAGAAA
Miro GTPase_2	Glyma03g30990	Transcribed	GCAGTTAGAAAGAATTCCTCTGGTTAGTCATC	GCCCAATACTGAACACGAATCCA	AGAAACTTTATACTGACAGAGCCGAC	GCAGTTAGAAAGAATTCCTCTGGTTAGTCATCAAGAAACTTTATACTGACAGAGCCGACTAAATTAACTGTTGGATTCGTGTTCAGTATTGGGC	95	75	GCAGTTAAAAAGAATTCCTCTGGTTAGTCATCAAGAAAATTTATAATGGCAGAGCTGACTTAAATTAACTGTTGAATTCGTGTTCAGTATTGGGC
SnRNP CorePro_2	Glyma03g30750	Transcribed	CCAGAAGGTGTGGACATTGAATTATTG	TGGAAGCCCACTGAATTCTCTCAT	ACGATGCCACAAGAAGAA	CCAGAAGGTGTGGACATTGAATTATTGCACGATGCCACAAGAAGAGAAGCCCGGTTAATGTCTTCATTTTGTGTTGTAATCTTATTTAATGAGAGAATTCAGTGGGCTTCCA	117	77	CCAGAAGGGGTGGACATTGAATTATTGCACGATGCCACAAGAAGAGAAGCCCGGGGCGGTTAGTGTCTTCATTTTGTGATGTATCTTATTTATGAGAATTCGGTGGGCTTACA

[a] Soybean sequence Glyma 1, Schmutz et al. (2010).

APPENDIX A

Chapter 4. Supplemental Table 2. Technical replicates for actin (Glyma19g00850), calculated mean, and relative expression calculated using y = 10^(mean sample Ct – b)/m), where b = slope and m = intercept from the standard curve validation (Applied Biosystems 2001).

		Cycle threshold (Ct)				
Sample Name	**Floral phenotypic class**	**Actin rep 1**	**Actin rep 2**	**Actin rep 3**	**Mean**	**Relative expression**
1-14-1	ms	28.8598	29.0681	29.0149	28.9809	9.7245
2-15-49	ms	28.4599	28.2003	28.3176	28.3259	15.0607
1-14-20	F	28.1984	27.9858	28.0835	28.0892	17.6396
3-14-8	F	26.7163	26.3320	26.2190	26.4224	53.6925
1-14-30	F	28.1380	27.9780	27.9489	28.0216	18.4545
3-14-20	ms	27.3373	27.0222	27.0943	27.1513	33.0013
1-14-44	F	27.8913	27.5798	27.6945	27.7219	22.5441
3-14-41	F	26.9967	26.7055	26.5510	26.7511	43.1115
1-14-55	ms	28.5449	28.2213	28.2853	28.3505	14.8150
3-15-1	F	27.1955	26.9685	26.9697	27.0446	35.4388
1-15-26	F	27.5235	27.3288	27.2299	27.3607	28.6928
3-15-15	ms	28.6892	28.3886	28.4228	28.5002	13.4060
1-15-56	ms	29.7819	29.4756	29.4415	29.5663	6.5779
3-15-31	ms	28.0526	27.8033	27.8332	27.8964	20.0641
2-14-19	F	27.6973	27.3811	27.5228	27.5337	25.5622
3-15-43	F	27.3822	27.0304	27.0649	27.1592	32.8278
2-14-27	ms	28.7352	28.6934	28.7009	28.7098	11.6545
4-14-4	ms	29.1567	28.6671	28.7561	28.8600	10.5425
2-14-51	F	27.6385	27.2061	27.4730	27.4392	27.2284
4-14-16	F	27.6096	27.0262	27.1777	27.2712	30.4619
2-15-9	F	28.4549	28.0881	28.1875	28.2435	15.9127
4-14-36	F	27.6617	27.1351	27.2708	27.3559	28.7865
2-15-36	F	28.2791	28.6667	28.3590	28.4349	14.0030
4-14-50	F	27.9207	27.4656	27.5134	27.6333	23.9185
4-15-4	F	26.7471	26.5535	26.5512	26.6172	47.1428
4-15-20	F	29.2244	28.6710	28.8457	28.9137	10.1711
4-15-52	ms	29.4430	28.9294	29.0081	29.1268	8.8217
1-14-3	ms	28.9407	28.7684	28.9808	28.8967	10.2875
2-15-51	F	27.7688	27.5964	27.3721	27.5791	24.7998
1-14-21	F	28.2410	28.0527	28.1095	28.1344	17.1156
3-14-11	ms	27.8003	27.6066	27.4592	27.6220	24.0990
1-14-35	ms	28.8396	28.5155	28.5356	28.6302	12.2910
3-14-22	ms	27.9190	27.5891	27.3534	27.6205	24.1233
1-14-45	ms	29.2929	29.1019	28.9539	29.1162	8.8844
3-14-43	F	27.2758	26.9617	27.0703	27.1026	34.0916

APPENDIX A

Chapter 4. Supplemental Table 2 (continued).

Sample Name	Floral phenotypic class	Cycle threshold (Ct)				Relative expression
		Actin rep 1	Actin rep 2	Actin rep 3	Mean	
1-15-11	F	29.2070	28.8780	28.9845	29.0232	9.4540
3-15-3	F	27.2907	27.0392	27.0126	27.1141	33.8294
1-15-29	F	28.4912	28.0566	28.2905	28.2794	15.5354
3-15-18	ms	28.8404	28.5599	28.5029	28.6344	12.2566
2-14-1	F	27.9828	27.4017	27.5814	27.6553	23.5687
3-15-32	ms	28.1921	27.9404	27.9570	28.0298	18.3537
2-14-20	ms	28.9700	28.7705	28.6853	28.8086	10.9104
3-15-46	F	28.0294	27.8908	27.8030	27.9077	19.9126
2-14-30	ms	28.4456	28.2239	28.0459	28.2384	15.9666
4-14-5	F	27.6555	27.5405	27.5241	27.5734	24.8947
2-14-53	F	27.9486	27.5277	27.6257	27.7007	22.8655
4-14-19	ms	28.1480	27.9634	27.9245	28.0120	18.5739
2-15-13	F	28.4512	28.1004	28.0295	28.1937	16.4510
4-14-37	ms	28.5448	28.1384	28.0616	28.2483	15.8620
2-15-37	F	28.0899	27.9630	27.9523	28.0017	18.7011
4-14-51	F	27.7716	27.1968	27.2334	27.4006	27.9388
4-15-11	F	27.1711	26.6448	26.6722	26.8294	40.9157
4-15-22	ms	33.0399	32.1543	32.3880	32.5274	0.9105
4-15-53	F	29.1768	28.4948	28.4086	28.6934	11.7832
1-14-5	ms	28.7190	28.7511	28.7152	28.7285	11.5105
2-15-53	F	27.6250	27.4953	27.3840	27.5014	26.1202
1-14-22	ms	28.8213	28.1788	28.3939	28.4647	13.7277
3-14-12	F	26.9123	26.7228	26.8305	26.8219	41.1205
1-14-36	ms	29.5959	29.2170	29.3194	29.3774	7.4623
3-14-23	F	27.6066	27.2411	27.3145	27.3874	28.1862
1-14-46	F	28.7406	28.4433	28.3396	28.5078	13.3376
3-14-45	ms	27.4510	27.2878	27.2781	27.3390	29.1129
1-15-13	F	28.8424	28.4208	28.5859	28.6164	12.4050
3-15-5	F	27.4816	27.1032	27.2468	27.2772	30.3392
1-15-34	F	28.0622	27.9504	27.8310	27.9479	19.3861
3-15-19	ms	28.3050	27.9978	27.9676	28.0901	17.6289
2-14-8	F	27.5440	27.2872	27.3256	27.3856	28.2213
3-15-33	F	27.7164	27.5619	27.5490	27.6091	24.3073
2-14-21	ms	28.5525	28.2971	28.3000	28.3832	14.4956
3-15-48	ms	28.0057	27.9073	27.9543	27.9558	19.2841
2-14-35	F	26.9955	26.7514	26.8162	26.8543	40.2391
4-14-6	ms	28.0698	27.8265	27.8253	27.9072	19.9196

APPENDIX A

Chapter 4. Supplemental Table 2 (continued).

Sample Name	Floral phenotypic class	Cycle threshold (Ct)				Relative expression
		Actin rep 1	Actin rep 2	Actin rep 3	Mean	
2-14-54	F	27.6495	27.1579	27.2842	27.3639	28.6325
4-14-22	ms	28.5255	28.2043	28.3748	28.3682	14.6413
2-15-17	F	27.9036	27.7288	27.6061	27.7462	22.1816
4-14-40	F	29.0295	28.4452	28.5191	28.6646	12.0120
2-15-38	ms	28.8357	28.5157	28.6168	28.6561	12.0806
4-14-52	ms	28.1099	27.7894	28.1341	28.0111	18.5842
4-15-12	ms	28.6174	28.1476	28.3015	28.3555	14.7661
4-15-26	F	27.9831	27.5981	27.4143	27.6652	23.4143
4-15-55	ms	29.3451	28.9104	28.7625	29.0060	9.5631
1-14-6	F	27.5170	27.6078	27.6551	27.5933	24.5656
2-15-55	F	27.6323	27.1731	27.4635	27.4230	27.5249
1-14-23	ms	28.6415	28.4953	28.5205	28.5524	12.9462
3-14-13	F	26.9830	26.6623	26.8556	26.8336	40.7991
1-14-37	ms	29.5004	29.2840	29.4023	29.3956	7.3724
3-14-25	ms	28.4549	28.1364	28.3096	28.3003	15.3204
1-14-47	F	27.8606	27.5556	27.6356	27.6840	23.1224
3-14-46	ms	27.9726	27.5390	27.8024	27.7714	21.8113
1-15-14	ms	29.3933	29.1209	29.5334	29.3492	7.6043
3-15-7	F	27.3361	26.9172	27.1421	27.1318	33.4328
1-15-40	ms	29.0335	28.8613	28.9301	28.9416	9.9832
3-15-20	F	27.6963	27.2474	27.3760	27.4399	27.2152
2-14-9	F	27.5587	27.6055	27.5984	27.5876	24.6598
3-15-35	F	27.8027	27.4516	27.2986	27.5176	25.8391
2-14-22	F	27.5969	27.1977	27.3882	27.3943	28.0575
3-15-49	F	27.4750	27.2304	27.3482	27.3512	28.8764
2-14-39	F	27.7795	27.7650	27.7044	27.7496	22.1302
4-14-7	F	27.8366	27.4793	27.5291	27.6150	24.2119
2-14-56	F	27.5592	27.3643	27.5668	27.4967	26.2019
4-14-25	ms	28.2222	28.0497	28.0679	28.1132	17.3591
2-15-23	F	28.0842	27.7697	27.8989	27.9176	19.7821
4-14-41	F	27.9522	27.5821	27.5525	27.6956	22.9432
2-15-39	F	28.2523	27.9566	28.0954	28.1014	17.4966
4-14-54	F	27.5540	27.3393	27.1951	27.3628	28.6531
4-15-13	F	27.0894	26.9224	26.9362	26.9826	36.9347
4-15-28	F	29.0964	28.2502	28.5289	28.6252	12.3324
4-15-56	ms	29.1696	28.7971	28.7832	28.9166	10.1511
1-14-8	ms	28.2138	28.2891	28.2194	28.2408	15.9418

APPENDIX A

Chapter 4. Supplemental Table 2 (continued).

Sample Name	Floral phenotypic class	Cycle threshold (Ct)				Relative expression
		Actin rep 1	Actin rep 2	Actin rep 3	Mean	
3-14-4	F	26.7908	26.7692	26.7716	26.7772	42.3659
1-14-24	ms	28.5187	28.4779	28.4286	28.4751	13.6327
3-14-14	F	26.7601	26.7025	26.4417	26.6348	46.5942
1-14-38	F	27.6341	27.4155	27.4818	27.5105	25.9628
3-14-26	F	26.9502	26.5704	26.6631	26.7279	43.7847
1-14-48	F	28.0489	27.8087	27.9092	27.9223	19.7203
3-14-49	ms	27.5987	27.1975	27.2343	27.3435	29.0249
1-15-15	ms	28.7765	28.8340	28.5837	28.7314	11.4876
3-15-9	ms	28.5628	28.1409	28.3790	28.3609	14.7129
1-15-44	ms	29.3096	28.9681	28.9477	29.0751	9.1316
3-15-22	ms	28.0267	27.6584	27.5445	27.7432	22.2252
2-14-10	ms	28.7522	28.6397	28.5193	28.6371	12.2345
3-15-37	ms	28.1882	27.9468	27.8758	28.0036	18.6778
2-14-23	F	27.4971	27.2307	27.2483	27.3254	29.3789
3-15-52	F	27.3625	27.0715	27.1577	27.1972	32.0041
2-14-42	ms	27.9362	27.6852	27.8099	27.8104	21.2500
4-14-9	F	27.3754	27.2541	27.1193	27.2496	30.9037
2-15-1	ms	28.9451	28.6695	28.6816	28.7654	11.2299
4-14-27	ms	28.0790	27.8216	27.9109	27.9372	19.5250
2-15-25	F	28.0677	27.5666	27.8814	27.8386	20.8538
4-14-42	ms	28.1864	27.9777	27.8312	27.9985	18.7420
2-15-42	F	28.0705	27.8541	27.8608	27.9285	19.6387
4-14-56	F	27.6315	27.4110	27.5291	27.5239	25.7316
4-15-14	ms	27.5717	27.4936	27.4643	27.5098	25.9734
4-15-30	F	27.9019	27.6439	27.5665	27.7041	22.8133
1-14-12	ms	28.5374	28.2679	28.4630	28.4228	14.1172
3-14-5	F	27.0283	26.6410	26.6555	26.7749	42.4300
1-14-26	ms	28.9183	28.6136	28.8409	28.7909	11.0400
3-14-16	F	27.0989	26.6543	26.7785	26.8439	40.5200
1-14-39	F	27.8612	27.4993	27.6990	27.6865	23.0828
3-14-27	ms	27.5718	27.3402	27.3263	27.4128	27.7128
1-14-50	F	27.8221	27.5896	27.6233	27.6783	23.2097
3-14-51	F	26.9886	26.7612	26.8597	26.8698	39.8243
1-15-21	ms	29.4566	28.9382	28.9355	29.1101	8.9206
3-15-10	ms	27.8791	27.5855	27.6296	27.6981	22.9053
1-15-45	F	28.2458	27.8227	27.9426	28.0037	18.6765
3-15-23	F	27.7778	27.5577	27.5594	27.6316	23.9446

APPENDIX A

Chapter 4. Supplemental Table 2 (continued).

Sample Name	Floral phenotypic class	Cycle threshold (Ct)				Relative expression
		Actin rep 1	Actin rep 2	Actin rep 3	Mean	
2-14-11	F	27.6443	27.4421	27.5627	27.5497	25.2910
3-15-38	F	27.1876	26.9393	27.1423	27.0897	34.3860
2-14-24	ms	28.2099	28.1962	28.1551	28.1871	16.5239
3-15-54	F	27.6090	27.3357	27.6143	27.5197	25.8039
2-14-45	F	27.3500	27.1599	27.1516	27.2205	31.5101
4-14-12	F	27.5202	27.1915	27.2065	27.3061	29.7597
2-15-2	F	28.4768	28.0325	28.0916	28.2003	16.3789
4-14-28	ms	28.3399	28.6152	28.0097	28.3216	15.1040
2-15-30	F	28.3037	27.9981	28.0327	28.1115	17.3792
4-14-46	F	27.5604	27.1935	27.1491	27.3010	29.8606
2-15-43	ms	29.0165	28.8668	28.8496	28.9110	10.1895
4-15-2	F	27.3224	26.9771	27.0557	27.1184	33.7340
4-15-15	F	26.9500	26.1008	26.6039	26.5515	49.2575
4-15-32	F	29.2354	28.5669	28.7022	28.8349	10.7210
1-14-18	F	27.3953	27.2928	27.2429	27.3103	29.6756
3-14-6	ms	27.7316	27.3344	27.4887	27.5182	25.8282
1-14-28	F	28.3569	28.1260	28.1747	28.2192	16.1733
3-14-17	ms	27.8337	27.3827	27.4672	27.5612	25.0980
1-14-41	F	28.1590	27.9042	27.9139	27.9924	18.8184
3-14-35	ms	27.6241	26.9426	27.1889	27.2519	30.8570
1-14-52	F	27.9657	27.6480	27.7459	27.7865	21.5912
3-14-53	ms	27.5372	27.1317	27.2857	27.3182	29.5199
1-15-24	F	27.9322	27.6044	27.8544	27.7970	21.4411
3-15-13	ms	27.8891	27.3414	27.5266	27.5857	24.6905
1-15-49	ms	28.8166	28.3479	28.7224	28.6290	12.3011
3-15-25	ms	28.2933	27.6641	27.9544	27.9706	19.0940
2-14-13	ms	28.1619	27.9839	28.1758	28.1072	17.4291
3-15-40	ms	28.2397	28.0266	28.0172	28.0945	17.5776
2-14-25	ms	27.8890	27.8288	27.6328	27.7835	21.6349
4-14-1	ms	28.0957	27.6754	28.0378	27.9363	19.5367
2-14-46	F	27.5772	27.2476	27.2442	27.3563	28.7772
4-14-13	ms	28.1228	27.8110	27.6989	27.8776	20.3176
2-15-5	F	28.0845	27.7831	27.6445	27.8374	20.8703
4-14-29	ms	28.0250	27.7575	27.9391	27.9072	19.9200
2-15-33	F	28.0845	27.8354	27.9332	27.9510	19.3450
4-14-48	F	27.5784	27.3496	27.3232	27.4171	27.6332
2-15-44	F	28.4445	28.2117	28.3104	28.3222	15.0980

APPENDIX A

Chapter 4. Supplemental Table 2 (continued).

Sample Name	Floral phenotypic class	Cycle threshold (Ct)				Relative expression
		Actin rep 1	Actin rep 2	Actin rep 3	Mean	
4-15-6	F	27.6367	27.3133	27.2990	27.4163	27.6476
4-15-18	ms	28.7095	28.2497	28.3693	28.4429	13.9292
4-15-33	F	27.6019	26.9895	27.1179	27.2364	31.1769
1-14-19	ms	28.8052	28.7043	28.6978	28.7358	11.4544
3-14-7	ms	28.3041	27.5476	27.8322	27.8947	20.0872
1-14-29	ms	28.4093	28.4002	28.3559	28.3885	14.4444
3-14-18	F	27.3561	26.7591	26.9785	27.0312	35.7559
1-14-43	ms	28.4531	28.3139	28.4896	28.4188	14.1543
3-14-37	ms	27.7560	27.3852	27.4291	27.5234	25.7389
1-14-54	ms	28.5964	28.3401	28.5474	28.4946	13.4557
3-14-54	F	26.9502	26.8587	27.0178	26.9422	37.9450
1-15-25	F	27.9433	27.5324	27.6427	27.7061	22.7828
3-15-14	ms	27.7606	27.2271	27.3201	27.4359	27.2877
1-15-53	ms	29.2042	29.0462	28.9853	29.0786	9.1107
3-15-27	F	27.6322	27.3712	27.4499	27.4845	26.4174
2-14-14	ms	28.0268	27.9896	27.9712	27.9959	18.7743
3-15-41	ms	28.1871	27.8906	27.8469	27.9749	19.0398
2-14-26	ms	28.0765	27.8503	27.6912	27.8727	20.3843
4-14-2	ms	28.0487	27.7315	27.7116	27.8306	20.9648
2-14-49	ms	28.0940	27.6697	27.8700	27.8779	20.3130
4-14-14	F	27.3464	27.0296	26.9992	27.1251	33.5832
2-15-6	F	27.6448	27.2742	27.3499	27.4229	27.5254
4-14-33	F	27.6203	26.9270	27.2075	27.2516	30.8620
2-15-35	F	27.8664	27.4274	27.4345	27.5761	24.8493
4-14-49	ms	27.7895	27.4320	27.5127	27.5781	24.8168
2-15-47	ms	29.9811	30.3832	29.9819	30.1154	4.5586
4-15-7	ms	28.6191	28.2730	28.2384	28.3768	14.5569
4-15-19	F	27.3172	27.0723	26.9629	27.1174	33.7549
4-15-34	ms	28.5721	27.8340	28.0333	28.1464	16.9784

APPENDIX A

Chapter 4. Supplemental Table 3. Fertile floral samples were normalized to fertile actin expression and male-sterile floral samples were normalized to male-sterile actin expression, then mean relative expression was calculated using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
1-14-18	F	62	4119.275	2927.659	20.862	24.753	1646.860	983.513	348.613	752.004	150.971
1-14-20	F	53	7021.589	4384.334	48.979	44.627	2462.630	980.470	640.393	755.209	158.830
1-14-21	F	25	6255.840	5119.147	98.579	43.283	2677.579	716.766	607.134	1111.400	253.691
1-14-28	F	37	9115.496	6584.899	10.109	166.060	3406.218	5499.063	418.078	1473.471	353.168
1-14-30	F	44	4642.061	3427.277	35.217	22.440	2109.111	1285.253	389.470	718.317	162.449
1-14-38	F	23	5009.532	3191.046	42.055	38.759	1935.033	846.443	448.630	722.026	170.948
1-14-39	F	42	5334.651	3209.866	53.076	29.328	2076.638	1900.324	558.711	805.320	177.665
1-14-41	F	65	6740.866	4366.984	14.553	47.111	2803.604	2585.742	754.687	919.130	211.532
1-14-44	F	65	5617.312	2518.670	43.753	21.651	2066.209	609.357	426.961	677.554	163.096
1-14-46	F	50	5181.235	4980.473	32.474	65.350	3784.900	859.092	500.744	1407.925	300.316
1-14-47	F	84	5824.141	3890.474	46.532	71.686	2810.689	738.617	325.990	1202.810	224.477
1-14-48	F	55	5506.142	3359.821	59.950	34.205	2805.719	903.545	516.805	1136.924	259.077
1-14-50	F	29	3698.434	2881.774	71.583	19.971	1915.641	956.307	484.944	771.368	188.922
1-14-52	F	44	5496.217	4614.163	66.004	28.166	2468.536	674.909	647.219	1121.883	261.242
1-14-6	F	31	3669.533	4037.867	52.033	18.214	1399.669	1175.895	332.812	737.348	190.473
1-15-11	F	32	15861.762	10883.133	155.154	50.645	4383.608	932.674	694.164	1491.672	490.543
1-15-13	F	22	11468.695	7576.068	320.259	48.364	4285.075	3579.394	900.581	1835.385	518.220
1-15-24	F	64	6569.115	8489.144	314.339	26.799	2208.611	2539.577	468.356	1129.391	315.576
1-15-25	F	38	5154.919	4231.585	60.503	43.368	2348.620	2479.704	555.497	754.004	155.723
1-15-26	F	35	5820.185	3035.390	44.398	25.727	1567.532	1395.633	468.186	574.483	119.275
1-15-29	F	59	7474.934	8921.023	201.771	71.799	2708.755	2953.971	425.246	1278.129	335.115
1-15-34	F	51	5621.976	11155.010	225.089	40.615	1986.054	1891.635	401.763	1145.855	294.802
1-15-45	F	24	8507.653	7712.688	125.270	52.189	3249.539	1768.187	608.844	1312.796	324.033
2-14-1	F	28	6388.550	3839.432	103.271	20.547	1627.107	2210.399	628.409	825.031	193.290
2-14-11	F	63	7113.644	4083.516	387.924	100.237	2308.200	1671.605	699.593	1094.599	277.221
2-14-19	F	129	5250.161	4398.490	117.355	28.693	1532.774	1102.607	502.366	820.076	211.781
2-14-22	F	38	4259.355	4102.538	102.839	29.631	1559.982	1078.022	492.369	849.259	240.718
2-14-23	F	83	7214.602	3938.055	91.081	44.527	2056.628	1946.989	653.687	995.782	215.221
2-14-35	F	145	4754.401	3205.713	75.818	20.673	1426.647	1915.160	468.352	753.504	194.098
2-14-39	F	103	10806.809	4172.392	77.189	25.222	2782.345	1470.799	946.617	1365.106	304.612
2-14-45	F	154	7035.451	4830.598	127.921	22.881	1652.164	1352.945	545.963	982.212	242.191
2-14-46	F	110	6344.399	3858.929	81.004	33.048	1836.003	3421.123	617.106	791.861	227.224
2-14-51	F	48	4690.428	3242.340	127.812	13.775	1133.654	1132.195	325.000	458.164	122.446
2-14-53	F	181	6342.050	4300.300	125.278	34.836	2456.386	1409.228	679.058	1112.823	244.125
2-14-54	F	185	7605.023	5079.212	81.728	33.096	2040.041	2992.085	561.778	986.395	229.745
2-14-56	F	62	6432.392	3448.149	71.819	16.849	1745.059	3092.319	570.563	914.304	220.308

APPENDIX A

Chapter 4. Supplemental Table 3. (continued)

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
2-14-8	F	100	5633.760	5269.340	170.692	33.987	1746.389	1542.615	415.701	941.575	222.788
2-14-9	F	161	7987.509	6046.425	108.375	21.847	2266.712	2773.156	606.247	1265.823	325.429
2-15-13	F	38	7779.734	4062.566	65.847	43.362	3122.076	520.509	535.608	1264.103	206.346
2-15-17	F	58	8677.329	4250.613	117.479	27.579	2963.493	1402.610	755.455	1345.341	355.679
2-15-2	F	98	13360.373	4976.611	105.443	39.063	3803.013	1756.540	579.814	1734.551	333.502
2-15-23	F	58	10752.237	4955.542	67.635	38.136	3015.943	1375.765	561.105	1272.968	280.904
2-15-25	F	62	12223.460	7512.053	35.372	26.107	3464.904	2350.108	501.417	1491.369	286.400
2-15-30	F	59	9275.657	5020.827	58.971	80.963	3503.323	2640.581	402.168	2141.659	353.521
2-15-33	F	49	10918.570	7522.818	67.807	33.866	3354.908	3741.613	647.537	1711.984	405.352
2-15-36	F	32	9121.788	5415.107	59.052	30.921	2656.380	1452.561	505.359	938.663	228.274
2-15-37	F	45	8817.193	4605.947	37.329	33.902	3202.876	1829.950	632.763	1414.034	251.620
2-15-39	F	70	10718.913	5603.402	62.921	35.851	3647.494	2780.779	446.382	1728.387	297.096
2-15-42	F	64	15223.250	4800.274	67.079	57.370	3782.015	1846.670	503.532	1769.667	307.582
2-15-44	F	66	12111.996	7211.054	81.657	42.347	3802.847	2433.068	635.627	2091.240	394.816
2-15-5	F	40	9259.158	5675.343	90.913	42.888	3114.544	719.202	640.843	1293.078	313.878
2-15-51	F	53	16086.851	10018.357	137.056	47.267	3459.106	1416.725	1107.219	1818.638	621.620
2-15-53	F	76	16904.873	8918.515	128.561	64.181	3169.013	1371.104	1159.896	2544.375	720.992
2-15-55	F	68	19664.313	8903.870	161.470	77.113	3641.480	1070.627	1237.996	2465.543	684.180
2-15-6	F	56	5999.007	3661.467	43.580	13.222	1939.321	813.021	362.017	771.539	185.481
2-15-9	F	50	8257.056	5234.184	40.370	29.134	2990.737	867.949	653.842	1123.944	223.025
3-14-12	F	64	5440.320	10437.458	185.561	24.428	1059.386	798.725	872.526	1218.015	383.536
3-14-13	F	87	9651.992	8422.636	179.572	26.597	1344.834	949.431	1126.966	1360.533	373.689
3-14-14	F	40	10553.960	6097.900	126.676	17.500	1268.493	990.163	1395.747	1062.143	259.793
3-14-16	F	85	8840.222	10992.985	166.371	35.564	1361.875	831.700	1036.263	1168.495	298.276
3-14-18	F	74	8224.285	12188.834	148.008	46.036	1660.652	1561.722	1175.506	1098.330	371.114
3-14-23	F	26	14520.370	10091.239	169.720	44.378	2360.538	1896.809	2164.911	2210.575	509.199
3-14-26	F	37	9010.657	7525.409	134.874	17.542	1393.661	1105.396	1407.649	1206.998	294.306
3-14-4	F	82	8305.376	9047.674	135.217	21.504	1300.296	1288.061	906.349	1037.139	278.685
3-14-41	F	50	7901.186	6150.365	61.608	25.359	1908.736	1340.754	1239.421	1053.999	246.408
3-14-43	F	101	11558.474	6515.237	81.529	27.623	1900.293	1086.012	1319.125	1366.869	345.935
3-14-5	F	46	10072.109	13041.625	190.573	50.100	1928.491	904.431	1407.303	1479.972	440.238
3-14-51	F	60	7949.344	6403.079	118.051	17.562	1482.877	874.133	780.469	1351.105	306.050
3-14-54	F	72	11327.082	8386.752	70.046	23.300	1615.465	2103.262	1444.251	2024.399	467.244
3-14-8	F	53	7553.083	7806.516	130.089	28.073	1194.240	770.408	1044.577	1201.738	351.401
3-15-1	F	46	8202.467	7707.404	68.777	25.895	1317.513	1084.467	958.098	1134.401	319.398
3-15-20	F	25	13034.408	7745.682	160.995	53.614	2184.680	639.201	1027.505	1348.001	302.762
3-15-23	F	29	11903.124	8476.095	156.212	76.196	2258.462	1116.121	886.666	1480.304	367.712
3-15-27	F	21	9063.398	5987.024	97.129	101.936	2412.782	1284.754	603.687	998.409	247.165
3-15-3	F	22	9951.754	6172.903	107.926	38.134	2015.391	944.168	1025.299	1181.074	291.023

APPENDIX A

Chapter 4. Supplemental Table 3. (continued)

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
3-15-33	F	72	14582.228	7512.427	161.850	35.347	2291.716	861.101	1189.001	1587.196	317.886
3-15-35	F	43	15558.524	7772.562	126.678	37.126	2417.260	845.874	1432.541	1700.293	418.899
3-15-38	F	75	12760.063	7092.025	184.121	46.041	2083.740	953.076	810.934	1596.281	370.762
3-15-43	F	81	12155.108	6359.319	137.848	37.923	2120.657	891.206	1097.414	1350.262	347.557
3-15-46	F	75	16847.931	9023.415	127.192	51.449	2385.088	765.196	1409.859	1633.992	519.122
3-15-49	F	48	15167.624	6140.924	193.103	25.453	1620.761	575.410	1090.381	1338.679	430.051
3-15-5	F	63	10376.700	7225.888	81.974	19.921	1859.737	1070.899	1195.934	1143.633	293.009
3-15-52	F	32	7302.844	7359.258	61.346	31.688	1502.653	576.896	622.269	855.963	245.690
3-15-54	F	13	11048.942	8159.773	41.718	45.481	1823.229	459.234	755.760	1380.964	363.967
3-15-7	F	57	10077.354	10960.221	129.703	42.703	1630.633	1141.577	903.719	1421.832	353.900
4-14-12	F	39	12100.097	6837.511	84.768	29.474	2068.452	1021.288	1177.200	1558.018	374.593
4-14-14	F	53	9432.474	5728.958	91.851	16.037	1732.668	1210.009	867.570	1059.320	292.574
4-14-16	F	46	13650.071	7769.538	59.331	61.652	2600.911	773.506	825.519	1515.287	302.409
4-14-33	F	32	14908.724	4904.484	88.272	55.032	2580.616	1170.058	854.639	1113.647	269.311
4-14-36	F	51	15124.513	7676.255	47.495	59.591	2588.918	729.065	957.476	1477.846	330.942
4-14-40	F	75	19334.414	14137.325	57.897	216.648	4415.919	1308.916	824.496	2712.252	651.806
4-14-41	F	65	15231.861	4291.056	62.610	77.356	3022.780	864.909	822.028	1683.528	277.826
4-14-46	F	26	12413.391	4573.008	50.263	71.905	2353.246	568.430	633.825	1554.027	286.534
4-14-48	F	41	10414.016	8095.648	45.451	53.171	2355.306	1012.029	570.453	1534.834	257.451
4-14-5	F	54	12054.001	9032.535	91.598	35.991	1838.464	793.365	863.662	1352.077	362.763
4-14-50	F	24	15059.068	7695.317	94.133	69.794	3367.876	897.989	1293.044	1571.646	338.818
4-14-51	F	37	10919.733	8102.974	49.791	57.163	2419.992	628.987	782.910	1212.465	261.186
4-14-54	F	49	11708.281	6357.420	85.721	28.471	2159.321	684.085	756.452	1139.379	271.896
4-14-56	F	36	15784.621	7240.043	34.709	88.689	2883.694	1178.999	744.316	1485.762	288.560
4-14-7	F	35	14698.217	8417.842	57.594	66.645	2381.922	1268.236	1004.890	1736.898	409.649
4-14-9	F	32	12151.081	8185.658	55.565	42.739	2106.131	678.406	1061.686	1437.918	357.029
4-15-11	F	82	14911.427	9046.775	507.981	56.509	3290.847	710.052	1393.084	2514.250	660.179
4-15-13	F	36	16576.146	13459.643	356.489	76.686	3416.120	1022.469	1222.289	3486.166	751.024
4-15-15	F	65	21734.013	12180.056	393.663	96.922	4156.901	1624.648	2148.858	3657.900	882.461
4-15-19	F	75	18154.189	16953.238	397.651	85.137	2664.459	2243.743	1341.278	2576.508	677.763
4-15-2	F	35	9163.874	6194.499	158.364	34.039	1928.291	1242.206	756.353	1099.359	325.358
4-15-20	F	45	13145.555	14132.414	378.912	58.645	2970.434	1545.458	697.470	1569.347	410.576
4-15-26	F	71	15035.800	10414.031	282.941	49.182	3425.560	1380.453	1212.590	2386.858	530.467
4-15-28	F	53	16654.530	11011.408	327.619	54.573	3297.470	1176.981	1327.305	2276.812	523.280
4-15-30	F	51	15231.253	11734.813	276.474	65.124	2986.790	2191.295	1021.341	2503.650	557.021
4-15-32	F	27	13799.011	9340.583	226.458	111.048	3323.076	1605.344	1180.553	2201.210	430.067
4-15-33	F	108	17746.138	8101.744	211.778	73.067	3537.291	1460.453	1298.145	2274.382	410.389
4-15-4	F	89	16883.823	8785.172	381.124	40.083	2813.058	504.599	1017.054	2008.294	491.226
4-15-53	F	59	26220.892	6964.609	241.042	60.303	3334.630	687.701	1218.071	1600.742	380.530

APPENDIX A

Chapter 4. Supplemental Table 3. (continued)

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
4-15-6	F	46	8907.675	7551.012	260.947	38.025	2615.921	1143.886	969.161	1297.393	239.951
2-15-35	F	80	7400.624	4139.066	44.729	21.628	2973.825	4347.699	551.251	1245.963	234.003
1-14-1	ms	12	5784.319	3797.925	44.560	20.695	1933.014	514.558	479.040	617.685	113.187
1-14-12	ms	21	5896.477	2902.027	62.044	100.102	2410.915	1503.607	404.432	897.867	182.532
1-14-19	ms	7	7767.492	4390.396	63.250	19.774	2645.574	1788.878	626.410	926.003	202.026
1-14-22	ms	19	4662.216	1833.466	50.162	29.839	1961.678	683.407	317.573	579.590	127.945
1-14-23	ms	16	7327.272	2277.121	41.739	31.672	2318.255	1281.011	657.995	899.376	184.190
1-14-24	ms	16	6420.321	2974.018	36.048	28.065	2010.435	2051.124	522.813	919.711	164.586
1-14-26	ms	9	5842.207	3609.578	50.793	22.663	2361.284	2444.085	452.877	911.031	232.327
1-14-29	ms	12	5466.196	3020.831	44.620	10.062	2051.444	1921.308	533.163	690.513	157.896
1-14-3	ms	13	4750.797	3553.689	58.345	44.618	2018.983	421.683	296.335	895.571	192.945
1-14-35	ms	10	5921.325	2447.909	40.039	34.999	2218.879	1976.647	646.597	843.024	191.395
1-14-36	ms	0	9211.772	3714.436	11.932	68.568	3615.568	3667.606	294.467	1386.194	280.579
1-14-37	ms	0	8549.783	3867.422	68.146	57.813	3334.704	895.252	483.092	1553.803	326.391
1-14-43	ms	7	6305.760	2576.924	51.844	13.877	1897.052	1589.069	445.030	778.025	171.001
1-14-45	ms	10	8015.619	2815.666	82.777	76.809	3336.563	1075.888	702.840	980.943	232.610
1-14-5	ms	13	4478.110	2816.139	45.470	34.204	1980.060	742.883	288.371	873.385	173.347
1-14-54	ms	11	7253.796	3097.778	69.304	41.039	2321.798	1500.133	675.101	892.269	189.697
1-14-55	ms	13	6818.475	2172.998	56.444	25.304	1941.808	567.929	488.849	731.944	131.195
1-14-8	ms	12	6079.049	3807.378	27.630	126.279	1959.904	3532.648	219.046	911.824	201.084
1-15-14	ms	0	12738.555	8514.907	170.293	43.118	3139.734	2448.286	532.024	1598.188	489.474
1-15-15	ms	1	8339.331	6028.718	196.578	42.103	2800.780	2393.870	623.922	1276.726	344.425
1-15-21	ms	2	10439.906	11844.170	246.756	57.438	3371.508	8173.816	712.805	1628.161	493.703
1-15-40	ms	0	9359.627	10353.050	146.458	48.009	2347.014	3346.926	447.093	905.981	295.878
1-15-44	ms	0	7579.109	9118.766	130.087	73.434	3986.624	3589.713	596.077	1360.843	365.320
1-15-49	ms	3	8001.072	8917.920	185.992	39.856	2245.145	1510.702	779.268	1035.943	340.032
1-15-53	ms	2	11461.132	8999.269	116.622	44.177	3088.187	6308.858	756.772	1437.067	460.400
1-15-56	ms	0	11035.053	8131.383	74.815	87.206	3821.248	2160.584	449.421	1265.363	328.651
2-14-10	ms	16	5823.576	8733.340	176.317	36.217	1624.480	3027.966	312.201	1083.663	258.979
2-14-13	ms	12	8214.189	6424.424	139.267	27.928	1872.133	1310.095	687.076	1072.009	263.544
2-14-14	ms	15	5588.212	4431.506	93.997	14.714	1670.451	2717.307	557.521	833.127	217.475
2-14-20	ms	14	5162.255	6116.126	101.327	18.606	1436.466	1185.382	263.875	889.897	230.657
2-14-21	ms	9	5507.876	4630.729	128.410	17.528	1613.365	1199.026	451.192	763.601	186.962
2-14-24	ms	21	7029.086	4139.777	121.559	16.067	1876.210	2211.780	645.642	1150.728	260.242
2-14-25	ms	29	6253.630	4464.876	88.928	29.509	1550.846	1832.445	497.932	910.618	213.953
2-14-26	ms	9	6006.830	4554.480	99.614	24.596	1554.721	2106.650	494.752	734.863	186.395
2-14-27	ms	0	7367.328	4848.019	188.895	19.841	1817.312	1641.652	523.716	841.646	174.626
2-14-30	ms	15	11640.246	4165.920	73.297	41.593	2820.725	1088.995	972.496	1161.046	252.868
2-14-42	ms	14	8566.318	3332.337	52.343	31.144	2127.673	1156.369	777.119	871.528	192.131

APPENDIX A

Chapter 4. Supplemental Table 3. (continued)

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
2-14-49	ms	36	6342.849	4684.590	113.194	32.068	1413.003	2090.397	482.627	731.199	174.713
2-15-1	ms	14	9929.651	4634.745	144.947	50.202	2779.639	1146.726	422.281	1506.256	334.575
2-15-38	ms	23	11783.619	3456.339	73.781	33.462	2519.283	1796.696	519.540	1084.701	228.325
2-15-43	ms	27	14909.893	7379.923	357.071	158.037	4181.394	2231.567	709.596	2482.749	433.478
2-15-47	ms	19	7822.305	14709.358	31.016	106.455	2880.662	661.405	335.971	1219.888	281.943
2-15-49	ms	19	22280.029	6300.487	211.252	59.775	3820.353	1687.202	1548.938	2665.723	710.465
3-14-11	ms	7	7708.047	9269.432	203.316	25.974	1295.345	686.899	1173.555	1324.710	392.632
3-14-17	ms	23	5433.841	12227.475	169.388	43.451	992.017	1043.588	681.379	1008.484	337.757
3-14-20	ms	27	10048.317	5536.032	111.681	22.753	1351.904	455.333	1114.672	997.070	229.656
3-14-22	ms	8	7885.172	9101.528	112.537	34.498	1319.546	846.481	1325.330	1195.891	347.004
3-14-25	ms	0	10829.727	12401.268	269.412	43.820	1973.192	1296.152	1046.398	2021.664	460.613
3-14-27	ms	5	7150.375	8200.142	116.972	22.830	1024.943	1010.429	752.870	1072.682	286.371
3-14-35	ms	12	6176.040	7146.535	65.446	35.056	1085.937	1286.861	841.028	829.751	224.172
3-14-37	ms	11	8729.721	8455.684	91.992	26.691	1528.752	1795.589	1225.277	1082.840	285.106
3-14-45	ms	29	8637.331	6507.047	64.093	24.162	1123.847	988.524	995.282	1020.270	253.618
3-14-46	ms	28	9266.661	8290.532	117.473	18.080	1465.190	1087.022	1045.914	1525.639	449.552
3-14-49	ms	39	5424.173	8134.325	121.590	26.687	1065.285	820.873	614.324	916.705	269.838
3-14-53	ms	34	7743.372	6135.734	99.919	17.482	1263.757	1243.813	860.671	1094.169	281.768
3-14-6	ms	13	6404.525	13597.450	192.533	27.505	1058.663	1039.524	892.832	1145.786	391.743
3-14-7	ms	0	7857.343	11734.605	260.870	23.693	1790.287	1941.687	1240.167	1316.811	370.794
3-15-10	ms	16	7269.734	5568.338	119.478	34.482	1384.504	837.856	678.923	1038.428	251.737
3-15-13	ms	33	10088.103	7231.385	114.529	35.382	1832.084	1336.549	1095.269	1066.538	249.296
3-15-14	ms	21	8794.295	5934.467	66.943	23.524	1723.311	1708.740	961.672	1062.508	255.535
3-15-15	ms	0	10627.668	8923.095	132.130	61.654	3088.272	1715.032	622.443	1762.322	361.994
3-15-18	ms	0	13178.510	9936.310	150.902	72.303	3496.595	1107.526	937.256	1876.385	468.341
3-15-19	ms	17	9037.232	7756.154	151.336	42.961	2211.514	682.927	590.383	1150.007	233.546
3-15-22	ms	38	8435.691	7378.748	83.876	25.071	1594.144	2142.579	860.982	984.856	220.371
3-15-25	ms	11	8571.033	9606.109	116.841	60.567	1962.387	895.873	768.302	1035.169	276.489
3-15-31	ms	16	10230.167	7555.135	127.482	26.636	1586.549	893.778	921.247	1046.941	265.662
3-15-32	ms	18	10220.928	7945.897	104.825	49.297	2014.266	933.747	950.700	1424.936	353.945
3-15-37	ms	25	11192.664	7382.579	123.447	24.376	1857.639	1106.249	1187.832	1246.482	298.600
3-15-40	ms	31	13217.626	8057.680	127.476	35.351	1864.055	849.536	927.278	1224.198	357.911
3-15-41	ms	35	11163.339	6871.887	152.261	35.013	1578.739	946.711	872.941	985.802	305.716
3-15-48	Ms	27	7809.743	5614.436	120.196	15.166	1545.524	448.164	883.481	1009.808	317.055
3-15-9	Ms	11	9152.331	5154.991	59.213	17.458	2145.441	3462.579	811.450	1158.166	292.267
4-14-1	Ms	28	11482.942	6053.521	75.520	28.679	1629.917	654.531	760.896	1296.532	309.063
4-14-13	Ms	23	11228.193	4368.033	68.809	34.484	1819.793	1034.459	743.774	1139.806	279.516
4-14-19	Ms	11	11959.543	7010.758	50.562	46.480	2205.458	621.649	732.437	1590.339	333.780
4-14-2	Ms	15	14549.649	3553.158	60.750	127.216	2110.306	928.726	1226.914	1393.338	297.997

APPENDIX A

Chapter 4. Supplemental Table 3. (continued)

Sample ID	Floral phenotypic Class	Seed count	snRNP (Glyma03g30880) mean	AUX/IAA responsive gene (Glyma03g31530) mean	Sucrose transporter (Glyma03g31020) mean	MYC (Glyma03g30940) mean	GTP-binding (Glyma03g31040) mean	Ca^{2+} transport protein (Glyma03g31420) mean	Isoamylase (Glyma03g30770) mean	snRNP core protein (Glyma03g30750) mean	Miro GTPase (Glyma03g30990) mean
4-14-22	Ms	0	12925.099	5411.976	83.366	65.377	2574.703	823.207	693.491	1452.732	327.634
4-14-25	Ms	13	10938.335	6430.107	55.065	46.012	1837.182	519.451	792.239	1313.867	342.511
4-14-27	Ms	22	9469.601	6250.026	60.505	23.022	1672.057	930.530	609.672	1218.274	291.971
4-14-28	Ms	14	13792.252	5003.670	75.296	68.722	2095.192	1337.215	648.999	1660.710	319.538
4-14-29	Ms	10	10961.688	4648.826	52.472	39.451	1685.400	838.265	603.098	1121.641	237.055
4-14-37	Ms	17	10700.786	7035.111	29.348	80.417	2563.969	838.980	793.388	1512.152	350.580
4-14-4	Ms	0	17636.256	8568.320	96.779	118.376	3843.004	1057.960	806.823	2543.694	612.035
4-14-42	Ms	26	9483.562	6552.593	62.482	29.765	1835.976	600.044	676.809	1311.991	300.422
4-14-49	Ms	14	8804.628	4088.074	50.362	28.526	1517.257	991.885	626.578	751.577	195.320
4-14-52	Ms	15	6891.300	7151.851	68.908	27.476	1588.917	710.219	586.246	912.219	245.930
4-14-6	Ms	19	9834.161	5263.531	65.271	43.145	1701.932	764.530	930.924	1023.447	280.932
4-15-12	Ms	2	16006.699	7061.206	293.090	38.687	2504.979	1051.412	1563.473	1939.563	435.664
4-15-14	Ms	2	7988.714	5813.025	272.627	32.367	2031.968	1712.455	1026.644	2018.394	464.400
4-15-18	Ms	0	12282.221	14741.204	324.551	57.938	2984.472	1202.391	945.419	2080.854	497.262
4-15-22	Ms	1	13915.095	13045.764	511.476	156.653	3109.403	1782.612	1281.712	1838.266	483.702
4-15-34	Ms	0	14264.063	9337.776	241.488	50.774	2806.758	1426.983	1167.814	1656.967	415.372
4-15-52	Ms	3	21999.565	6993.889	228.059	80.831	3280.048	852.786	1264.369	1980.671	420.140
4-15-55	Ms	4	18678.312	16231.591	203.894	134.536	5149.898	1825.379	1333.014	3505.060	814.768
4-15-56	Ms	0	23017.615	12857.547	403.951	57.477	4441.378	1437.740	1617.327	3631.363	895.897
4-15-7	Ms	2	11156.307	8015.518	217.190	29.631	2668.222	2311.928	1028.119	1167.801	409.108

APPENDIX A

Chapter 4. Supplemental Table 4. Using a split-plot experimental design in which whole-plots were temperature regimens across growth chamber and sub-plots were floral phenotypic classification within chambers (Wiebbecke et al. 2011a) an Analysis of Variance (ANOVA) analyzed by PROC MIXED in SAS version 9.2 (SAS Institute Inc. 2008) was used to estimate mean relative quantity and standard errors for the internal control (actin). Candidate genes were normalized by relative quantity of actin mRNA for their respective floral phenotypic classification. Relative gene expression of technical replicates was calculated using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001), averaged for each sample, and compared using the split-plot experimental design described above.

Putative gene annotation **Glyma 1 assembly ID**[c]	**Floral phenotypic Class**	**Mean relative expression**	**SE**	***P*-value**	**EnvID *P*-value**	**Pheno *P*-value**	**EnvID×Pheno *P*-value**
Actin	F	26.9078[a]	1.8496	0.035			
Glyma19g00850	ms	15.8242[a]	1.8614	0.0594	0.5405	<0.0001	0.1196
Isoamylase	F	9.5669[b]	0.1113	0.0042			
Glyma03g30770	ms	9.478[b]	0.1123	0.0036	0.4057	0.1326	0.5406
AUX/IAA responsive gene	F	12.656[b]	0.05347	<.0001			
Glyma03g31530	ms	12.6566[b]	0.05552	<.0001	0.1703	0.9913	0.0697
Ca^{2+} transport protein	F	10.3236[b]	0.2122	0.0093			
Glyma03g31420	ms	10.4938[b]	0.2136	0.0083	0.7094	0.0713	0.1953
GTP binding	F	11.1402[b]	0.1208	0.005			
Glyma03g31040	ms	11.0829[b]	0.1215	0.0046	0.5351	0.2389	0.2845
MYC	F	5.2839[b]	0.07021	<.0001			
Glyma03g30940	ms	5.2727[b]	0.07477	<.0001	<.0001	0.9132	0.7260
Miro GTPase	F	8.259[b]	0.0403	<.0001			
Glyma03g30990	ms	8.245[b]	0.04292	<.0001	<.0001	0.8125	0.2640
snRNP	F	13.1484[b]	0.04942	<.0001			
Glyma03g30880	ms	13.1525[b]	0.05131	<.0001	0.1776	0.9395	0.0327
snRNP core protein	F	10.3078[b]	0.04503	<.0001			
Glyma03g30750	ms	10.2557[b]	0.04724	<.0001	0.1778	0.3553	0.7716
Sucrose transporter	F	6.7779[b]	0.1109	0.0023			
Glyma03g31020	ms	6.8294[b]	0.1134	0.0014	0.2347	0.5812	0.9428

[a] Each samples relative quantity mRNA was calculated using y = 10^(sample Ct – b)/m), Applied Biosystems (2001).

[b] Each samples relative expression, $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) expressed as mean fold change on the $\log_2$ scale.

[c] Soybean sequence Glyma 1, Schmutz et al. (2010).

Chapter 4. Supplemental Table 5. Using a split-plot experimental design in which whole-plots were temperature regimens across growth chamber and sub-plots were floral phenotypic classification within chambers (Wiebbecke et al. 2011a) an Analysis of Variance (ANOVA) analyzed by PROC MIXED in SAS version 9.2 (SAS Institute Inc. 2008) was used to estimate mean relative quantity of actin or the relative candidate gene expression for each floral phenotypic classification within each environment.

Env ID	Night / Day (°C)	Floral Phenotypic Class	No. samples	Actin (Glyma19g 00850)[a]	Actin SE	Iso amylase (Glyma03g 30770)[b]	Iso amylase SE	AUX/IAA responsive gene (Glyma03g 31530)[b]	AUX/IAA responsive gene SE	Ca2+ transport protein (Glyma03g 31420)[b]	Ca^{2+} transport protein SE	GTP-binding (Glyma03g 31040)[b]	GTP-binding SE	MYC (Glyma03g 30940)[b]	MYC SE	Miro GTPase (Glyma03g 30990)[b]	Miro GTPase SE	snRNP (Glyma03g 30880)[b]	snRNP SE	snRNP core protein (Glyma03g 30750)[b]	snRNP core protein SE	Sucrose transporter (Glyma03g3 1020)[b]	Sucrose transporter SE
v214	12/ 30	F	15	27.488	5.05	9.141	0.30	12.033	0.14	10.818	0.58	10.839	0.33	4.825	0.18	7.822	0.10	12.633	0.13	9.842	0.12	6.792	0.30
214	12/ 30	ms	12	16.798	5.10	9.035	0.31	12.256	0.15	10.728	0.59	10.773	0.33	4.614	0.20	7.749	0.12	12.726	0.14	9.827	0.13	6.761	0.31
314	15/ 30	F	14	40.730	5.06	10.226	0.30	13.059	0.14	10.130	0.58	10.569	0.33	4.769	0.19	8.426	0.11	13.154	0.13	10.355	0.12	7.005	0.30
314	15/ 30	ms	14	25.810	5.06	9.910	0.30	13.093	0.14	10.029	0.58	10.324	0.33	4.756	0.19	8.317	0.11	12.901	0.13	10.171	0.12	7.018	0.30
115	17/ 35	F	8	18.547	5.22	9.093	0.32	12.806	0.17	10.988	0.60	11.389	0.34	5.418	0.25	8.171	0.14	12.917	0.16	10.136	0.15	7.216	0.33
115	17/ 35	ms	8	9.390	5.22	9.228	0.32	13.111	0.17	11.657	0.60	11.571	0.34	5.710	0.25	8.581	0.14	13.247	0.16	10.334	0.15	7.228	0.33
315	18/ 30	F	15	28.972	5.05	9.923	0.30	12.870	0.14	9.726	0.58	10.939	0.33	5.355	0.18	8.406	0.10	13.496	0.13	10.368	0.12	6.821	0.30
315	18/ 30	ms	15	19.147	5.05	9.743	0.30	12.827	0.14	10.130	0.58	10.913	0.33	5.081	0.18	8.203	0.10	13.257	0.13	10.205	0.12	6.819	0.30
415	18/ 35	F	13	27.986	5.06	10.178	0.30	13.291	0.14	10.257	0.58	11.589	0.33	5.925	0.19	8.942	0.11	13.913	0.13	11.059	0.12	8.234	0.30
415	18/ 35	ms	9	12.850	5.18	10.263	0.31	13.264	0.16	10.503	0.59	11.599	0.34	5.916	0.23	9.008	0.13	13.848	0.15	11.023	0.14	8.163	0.33
114	21/ 30	F	15	21.001	5.05	8.902	0.30	11.907	0.14	10.128	0.58	11.196	0.33	5.211	0.18	7.702	0.10	12.398	0.13	9.849	0.12	5.327	0.30
114	21/ 30	ms	18	12.070	5.02	8.813	0.30	11.559	0.14	10.351	0.58	11.168	0.33	5.143	0.16	7.537	0.09	12.632	0.13	9.781	0.11	5.551	0.29
215 & 414	23/ 30	F	27	23.631	3.55	9.505	0.21	12.626	0.10	10.219	0.41	11.461	0.23	5.485	0.12	8.343	0.07	13.528	0.09	10.545	0.08	6.051	0.21
215 & 414	23/ 30	ms	29	14.706	3.66	9.355	0.22	12.488	0.11	10.059	0.42	11.233	0.24	5.689	0.16	8.320	0.09	13.457	0.11	10.449	0.10	6.265	0.23

[a] Each samples relative quantity mRNA was calculated using y = 10^(sample Ct – b)/m), Applied Biosystems (2001).

[b] Each samples relative expression, $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001) expressed as mean fold change on the $\log_2$ scale.

[c] Soybean sequence Glyma 1, Schmutz et al. (2010).